Navy Personnel Research, Studies, and Technology Division
Bureau of Naval Personnel (NPRST/PERS-1)

Millington, TN 38055-1000

NPRST-TR-07-2 April 2007

RIDE vs. CLASP Comparison and Evaluation: Models and Parameters

John S. Folchi
Independent Contractor

NPRST
research at work

NPRST-TR-07-2
April 2007

RIDE vs. CLASP Comparison and Evaluation: Models and Parameters

John S. Folchi
Independent Contractor

Reviewed and Approved by
Jacqueline A. Mottern, Ph.D.
Institute for Selection and Classification

Released by
David L. Alderton, Ph.D.
Director

Navy Personnel Research, Studies, and Technology
Bureau of Naval Personnel
5720 Integrity Drive
Millington, TN 38055-1000
www.nprst.navy.mil

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
30-04-2007	Final Technical Report	Jul 2003 - Dec 2003

4. TITLE AND SUBTITLE RIDE vs. CLASP Comparison and Evaluation: Models and Parameters	**5a. CONTRACT NUMBER** DAAD19-02-D-0001
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) John S. Folchi	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) John S. Folchi (Independent Contractor) 2734 Tokalon St. San Diego, CA 92110	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Navy Personnel Research, Studies, and Technology 5720 Integrity Drive Millington, TN 38055-1000	**10. SPONSOR/MONITOR'S ACRONYM(S)** NPRST
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) NPRST-TR-07-2

12. DISTRIBUTION/AVAILABILITY STATEMENT

A - Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

CLASP, the Navy's enlisted personnel classification algorithm, has been used operationally since the early 1980s. The Navy is considering replacing it with the RIDE algorithm, as part of its transition from PRIDE to NRAMS. The purpose of this research was to evaluate and compare the algorithms, so classification policymakers could better judge the feasibility and wisdom of the proposed change. The RIDE algorithm was analyzed to identify and evaluate the assumptions upon which it is based. Particular attention was given to the First Pass Pipeline Success (FPPS) school performance criterion and the Point of Diminishing Returns (PDR) concept. RIDE school performance prediction models were compared with logistic regression prediction models to evaluate the RIDE mode development methodology. In addition, corresponding components of CLASP and RIDE were compared to better understand their similarities and differences. Sparsely documented aspects of the CLASP model were described in detail so policymakers could better understand the unique characteristics of CLASP and the methodology used to develop the CLASP utility functions.

15. SUBJECT TERMS

Personnel classification, Classification and Assignment within PRIDE (CLASP), Rating Identification Engine (RIDE), Armed Services Vocational Aptitude Battery (ASVAB), person-job matching, utility of person-job assignments

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Rhonda Patton
					19b. TELEPHONE NUMBER *(Include area code)*
UNCLAS	UNCLAS	UNCLAS	UNCLASS	75	901-874-4994 (882)

Foreword

This report compares the current enlisted job classification algorithm, Classification and Assignment within PRIDE (CLASP) instituted in 1981, with a proposed replacement algorithm, the Rating Identification Engine (RIDE). RIDE was developed over the course of several years, beginning with funding from the Office of Naval Research (Code 34, PE 0603236N), augmented by funding from Commander Navy Recruiting Command to accelerate its development. The motivation to build a replacement for CLASP was two-fold. First, components of CLASP are not well documented and it executes off an expensive mainframe computer system. Second, CLASP has a number of "hard coded" components that are inflexible and difficult to maintain. In contrast, RIDE is web-based and flexible. The flexibility to add new classification rules, filters, and tests was seen as an important component of our research program to overhaul and improve the Navy's enlisted selection and classification process.

RIDE substantially met the design requirements, it has an easy to use interface, can be reconfigured rapidly and easily, and most importantly, new tests or classification tools can be easily integrated. However, RIDE was under an accelerated development cycle to meet deadlines to coincide with a planned overhaul of the Navy's recruiting management system (of which CLASP was one component). As a result, RIDE was not as thoroughly evaluated against CLASP as would otherwise have been done. The current report provides a detailed evaluation of both CLASP and RIDE and compares them in terms of their embedded philosophies, functionality, maintenance, and efficacy. In the end, it is clear that the continued use of CLASP is indefensible for a number of reasons. Nevertheless, there are several concerns with RIDE that should be remedied and a plan is needed to refresh its parameters to maintain its integrity across time.

This specific work reported here was supported by the Navy Personnel Research, Studies, and Technology department (Ms. Janet Held) through the U.S. Research Office of Scientific Services Program administered by Battelle (Delivery Order 296, Contract No. DAAD19-02-D-0001).

David L. Alderton, Ph.D.
Director

Contents

List of Figures

Introduction

Across the United States, Military Entrance Processing Stations (MEPS) process approximately 51,000 applicants for enlistment into the active U.S. Navy each year. Of these, upwards of 39,000 are selected for enlistment (Kaemmerer, G., personal communication, September 14, 2006). Each of the selected applicants must be classified into a Navy rating (i.e., job) that matches the individual's abilities and is needed by the Service. To accomplish this, a job-matching algorithm is employed. The current Fortran-based algorithm was originally developed in the late 1970s and put into wide-scale Navy use in 1981. This paper compares and contrasts the existing classification software with a newly designed classification algorithm.

Rating Identification Engine (RIDE)

The Personalized Recruiting for Immediate and Delayed Entry (PRIDE) system is the Navy's current overarching computer system for processing applicants for enlistment into the Navy. The Rating Identification Engine (RIDE) is an enlisted Navy rating job classification algorithm that is designed to replace the Classification and Assignment within PRIDE (CLASP) algorithm. RIDE consists of two components: (1) the School Pipeline Success Utility (SPSU) and (2) the Armed Forces Qualification Test (AFQT). The two RIDE components are designed to work in close association with each other as opposed to the more or less independent operation of the six CLASP components.

Classification and Assignment within PRIDE (CLASP)

Much of the material in this report is quoted directly from Kroeker and Rafacz (1983), which describes the five components of the original CLASP model implemented in 1981. Kroeker and Folchi (1984) describe the Attrition Component, which was added to CLASP in 1983.

The CLASP utility model was formulated to ensure consistent application of Navy personnel classification policy among classifiers and from one assignment to the next. It is comprised of six components: School Success, Aptitude/Complexity, Navy Priority/Individual Preference, Minority Fill, Fraction Fill, and Attrition. Each component was designed to influence a composite utility calculation independently of the others. This design does not imply strict statistical independence; rather, a slight degree of correlation among the utility components is expected. The magnitude of these correlations has never been studied.

The School Success, Aptitude/Complexity, Navy Priority/Individual Preference, and Attrition are often called "Fit" components, because they optimize job assignments based upon psychologically-based goodness-of-fit measures. The Aptitude/Difficulty, Priority/Preference, and Attrition components are very similar because they are based on policymaker judgments concerning the value to the Navy of assigning an individual with a given person attribute to a job with a given job attribute. The school success component differs from the other "Fit" components because its utility model is based upon the empirical relationship between “A” School performance and Armed Services Vocational Aptitude Battery (ASVAB) composite scores.

The Minority Fill and Fraction Fill components are often referred to as "Fill" components, because they optimize based upon the goal of achieving approximately equal fill rates during each recruiting period. The Minority Fill component focuses upon achieving appropriate balance between minority and non-minority accessions in each job category, while the Fraction Fill component is focused on achieving uniform quota fill rates across job categories.

Using a utility function whose mathematical form is unique to it, each CLASP component computes the raw utility value of the prospective person to job assignment. Then, using mean and standard deviation parameters that describe the distribution of utility values in the reference population, each raw utility is standardized so that its mean is 50 and its standard deviation is 10.

Both the RIDE and CLASP algorithms can be conceptualized as operating on a payoff matrix, which is a rectangular array of numbers representing the utilities of the various decision outcome combinations. Assume that there are m individuals to be assigned to jobs and n job openings. If individuals are indexed by i ($1 \le i \le m$) and jobs are indexed by j ($1 \le j \le n$), then the entry $U_{i,j}$ in row i and column j of the matrix expresses the value to the Navy (on an arbitrary scale) of assigning the i^{th} person to the j^{th} job. Higher payoff values are more desirable than the lower ones, because Navy policy considers the probability of success on a job to be a monotonically increasing function of payoff value. The payoff matrix may be used for both comparisons across jobs and comparisons across individuals. Thus, $U_{i,j_1} > U_{i,j_2}$ implies that individual i is better suited for job j_1 than job j_2, while $U_{i_1,j} > U_{i_2,j}$ implies that individual i_1 is better suited for job j than individual i_2.

Ideally, the composite utility function for each job category should be a realistic mathematical representation of the value of assigning a given person to that job, based upon all identifiable factors considered relevant to the classification decision. However, because there are several important factors that neither RIDE nor CLASP are able to incorporate into the classification process, the goodness-of-fit measures they generate are often only a small part of the information factored into the final classification decision.

School Pipeline Success Utility (SPSU) Component of RIDE and School Success Component of CLASP Comparison

This section evaluates the empirical relationship between composite score and “A” School performance, and the manner in which that relationship is incorporated into CLASP and RIDE, with emphasis on RIDE. In particular, we want to know how well the applicant's ASVAB composite score predicts “A” School performance. We also evaluate the “Point of Diminishing Return(s)” (PDR) concept.

The PDR concept hypothesizes the following general relationship between composite score and school performance: The relationship is monotonically increasing at the lower end of the composite score distribution, including the region to the immediate right of the cut score. However, as the composite score increases toward the PDR, the rate of performance improvement declines and eventually flattens out at the PDR. Between the PDR and the high end of the score distribution, school performance either remains flat or declines. The leveling off or decline may be attributed to high aptitude students who are "over-qualified" for the curriculum/career path they are being considered for and, thus, may be better suited for a more challenging training curriculum and/or career path.

The following procedure (Folchi, 1999) was used to model the empirical relationship between composite score and First Pass Pipeline Success (FPPS) in each of 70 “A” School samples and determine the PDR in each sample. The data consisted of students enrolled in the “A” School training pipelines for 70 ratings during fiscal years 1996, 1997, and 1998. The primary ASVAB selector composite score and FPPS status were available for each student in each sample. The dichotomous criterion FPPS is coded as 1 (one, success) if the student completed all courses in his “A” School pipeline without any course failures or setbacks, and as 0 (zero, failure) otherwise. The procedure defines a methodology for grouping adjacent data points into groups (hereafter called "bins") that are (somewhat) evenly spaced along the composite score distribution.

Starting at the high end of the distribution, the procedure sequentially constructs bins by moving toward the low end in bin range increment of 5 points. The procedure adds all points in each increment to the bin, and continues on to the next increment, until a minimum bin size of 10 or more points have been added to the bin. After the bin membership has been determined in this manner, the bin is identified with a value on the composite score scale equal to the midpoint of the maximum and minimum of scores in all increments used to build the bin. Construction of the next bin (to the left of the bin just completed) starts at the point immediately to the left of the minimum score in the previous bin. The FPPS rate among students in the bin associates each bin with a point on the conditional probability of FPPS scale on the composite score. The PDR is found by determining all bins whose FPPS rates are within 1 percent of the bin having the largest FPPS rate. The PDR is the lowest composite score associated with the bins from this set. The bin in which the PDR is located is called the PDR bin and the bin in which the Cut Score is located in called the Cut Score bin. The associated points are named accordingly: (CS, F_{CS}) is the Cut Score point and (PDR, F_{PDR}) is the PDR point.

SPSU Component of RIDE

The SPSU component description is based on Folchi (1999). The description has been broken into 3 stages to provide a more detailed and understandable explanation.

Stage I: Bin FPPS Rate Model

The Stage I model is the result of the bin construction algorithm after all adjacent bins have been connected by line segments. Its equation is given in Appendix A. As shown in Figure 1, the Stage I model is piece-wise linear such that each segment provides a linear interpolation estimator of the conditional probability of FPPS for composite scores between the midpoints of adjacent bins. However, due to its complexity, the Stage I model was transformed into Stage II.

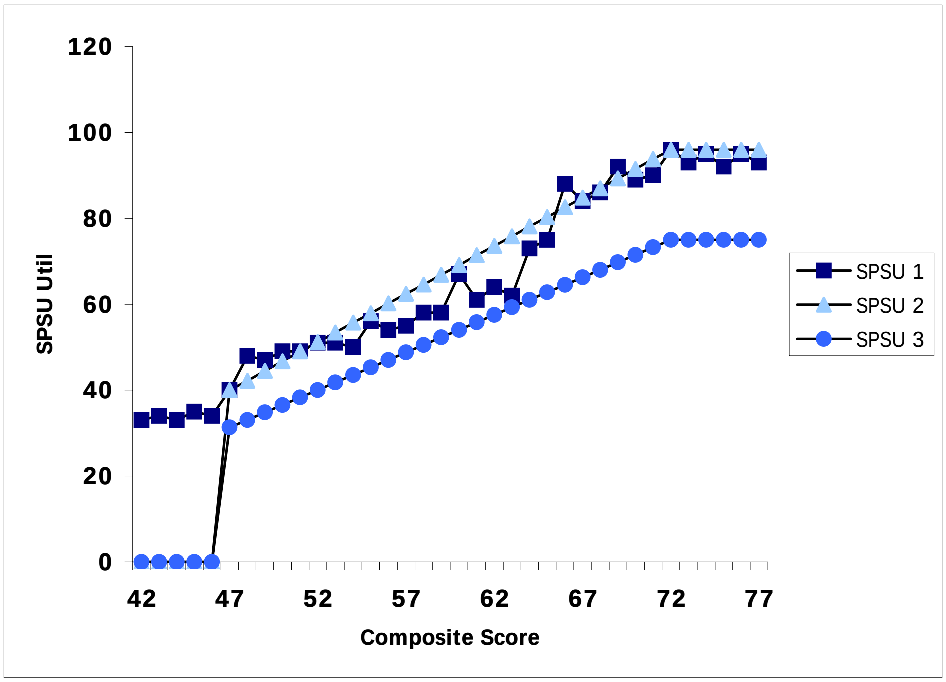

Figure 1. School Pipeline Success Utility

Stage II: Non-standardized FPPS Prediction Model with PDR

The Stage II model is constructed by eliminating all bins and line segments in the Stage I model, except the Cut Score (CS) and PDR points. The line segment between these 2 points estimates the conditional probability of FPPS for each composite score in the interval CS $\leq$ X $\leq$ PDR. For X < CS, the SPSU is zero, as defined by the horizontal line starting at X = CS - 1 and extending to the left toward the minimum composite score. For X > PDR, Stage II model is defined by the horizontal line starting at the PDR point and extending to the right toward the maximum composite score.

As shown in Figure 1, the Stage II model simplifies the Stage I model because it has no more than three line segments. It is considered unstandardized because it has not been adjusted so that meaningful comparisons across job options are possible. The Stage II equation is given by:

Regardless of whether $PDR_j = CS_j$ or $PDR_j \neq CS_j$:

$$S^{II}_{i,j} = 0 \quad \text{if} \quad X_{i,j} < CS_j,$$

$$S^{II}_{i,j} = 100 F_{PDR_j} \quad \text{if} \quad PDR_j \leq X_{i,j}. \qquad \text{Stage II Equation}$$

If $CS_j \leq X_{i,j} \leq PDR_j$ and $CS_j \neq PDR_j$:

$$S^{II}_{i,j} = 100\left\{\left(\frac{F_{PDR_j} - F_{CS_j}}{PDR_j - CS_j}\right)\left(X_{i,j} - CS_j\right) + F_{CS_j}\right\}.$$

where F_{PDR} and F_{CS} are the FPPS rates in the PDR and cut score bins, respectively,

CS_j = Cut Score for composite associated with job option *j*,

PDR_j = PDR for job option *j*, and

$X_{i,j}$ = ASVAB composite score for individual *i* in job option *j*.

The conditional FPPS probabilities provided by the Stage II model cannot be meaningfully compared across different job options. If applicants were assigned to jobs solely on the basis of their conditional FPPS probability, then most would be assigned to easy schools and few would be assigned to difficult schools, since the easier schools generally have larger FPPS probabilities. (Ease and difficulty in this context refer to both the proportion of the applicant population satisfying the ASVAB selection standard and the proportion of student population satisfying the FPPS criterion.) For example, suppose an applicant has the same conditional probability of FPPS in schools A and B and is qualified for both schools. Assume also that A uses a more stringent ASVAB selection criterion than B and that A graduates a smaller proportion of students than B. One may argue that it would be more beneficial to send this applicant to A than to B. Accordingly, the SPSU Stage III model adjusts the Stage II conditional probability of FPPS estimate for two measures of school difficulty: (a) difficulty experienced by the average applicant population member in satisfying the ASVAB qualification standard, and (b) difficulty experienced by the average "A" School qualified student in satisfying the FPPS criterion.

Another method of counteracting the tendency for school success utility scores to put too many applicants in easy schools is to design the remaining classification model components to compensate for this tendency. For example, the CLASP Aptitude/Difficulty component counteracts the CLASP School Success component in this respect. In RIDE, both the transition from Stage II to Stage III and the RIDE AFQT component fulfill the compensatory role.

The Hardness index is a measure of the difficulty that the average applicant population member experiences in satisfying the “A” School ASVAB qualification standard. It assumes values between zero and one, where zero indicates the minimum difficulty and one indicates maximum difficulty. The hardness index for job option j is defined as:

Let NRJO be the total number of RIDE job options.

Let NT_j be the number of ASVAB subtests in composite for job option j.

Let $H_{Max} = Max\left(\frac{CS_j}{NT_j} \mid 1 \le j \le NRJO \right)$, and

let $H_{Min} = Min\left(\frac{CS_j}{NT_j} \mid 1 \le j \le NRJO \right)$. The hardness factor is given by:

$$H_j = \frac{\frac{CS_j}{NT_j} - H_{Min}}{H_{Max} - H_{Min}} .$$

The adjustment for the difficulty experienced by the average student in satisfying the FPPS criterion is determined by the reciprocal of the FPPS rate at the PDR. This, of course, assumes that the FPPS rate at the PDR is representative of the FPPS rate of all students taking the course. The smaller the FPPS rate at the PDR, the larger the reciprocal is, and, therefore, the greater the difficulty of satisfying the FPPS criterion. Thus, the transition from Stage II to Stage III will produce a larger upward shift for a school in which it is more difficult to satisfy the FPPS criterion. The standardization factor is the ratio of the hardness index to the FPPS rate at the PDR. The larger the hardness index and the smaller the FPPS rate at the PDR, the larger the standardization factor. Accordingly, the Stage III model is

$$S_{i,j}^{III} = \frac{H_j}{F_{PDR_j}} S_{i,j}^{II} \qquad \text{Stage III Equation}$$

Observe from Figure 1 that the Stage II and Stage III models are discontinuous between cut score minus one and the cut score, unless the FPPS rate in the cut score bin is zero. Thus, there can be a large difference between the SPSU value at the cut score and the SPSU value (zero) everywhere below the cut score.

Criticisms of RIDE SPSU Model

Stage I Model, Bin Construction, and PDR Determination

Any procedure for grouping data in this manner is arbitrary. Application of different grouping procedures, bin sizes, and bin range increments lead to different bin memberships. Different bin memberships in turn produce different empirical relationships between FPPS and composite score and, consequently, different PDRs and different SPSU models. Furthermore, there is no a priori reason to believe that any one combination of grouping procedure, bin size, and bin range increment is superior to any other. Increasing Bin Size improves the accuracy of the FPPS rate estimates in each bin. However, it does so by reducing the number of bins and increasing the length of the interval between bins. As a result, the identification of each bin with a particular composite score becomes more arbitrary and diffuse. In addition, although there is no a priori reason to believe that the PDR necessarily exists, the PDR search procedure has been defined in such a manner that it will <u>always</u> find one.

FPPS School Performance Criterion

Several potential problems may arise as a consequence of using FPPS. To the author's knowledge, FPPS has never been studied or utilized in previous NPRDC/NPRST selection and classification research. It is not possible to anticipate how well it will perform in comparison to school performance measures used in ASVAB validation studies, such as final school grade (FSG).

A tailor-made school performance measure is usually developed during the course of performing an ASVAB validation study. Developing such a measure is often difficult and time-consuming because a detailed understanding of the course and student evaluation process is required. However, from the author's perspective, the effort generally produces a criterion that does well in differentiating students from one another. The resulting validity coefficients seem, in general, to be larger than those derived from more readily available performance measures, such as those obtained from Navy Integrated Training Resources and Administration System (NITRAS). A corollary to this observation is that differences between FPPS and a tailor-made performance criterion may be substantial enough to produce different validation study outcomes. For example, the ASVAB composite that correlates the highest with FPPS in a particular "A" School pipeline may not be the same as the composite that correlates highest with a criterion that is tailor-made for the "A" School in that pipeline.

This has important implications for RIDE. The SPSU model and parameters were developed using FPPS as the school performance measure and the current ASVAB selector composite as the student aptitude measure. However, no research has verified that the current ASVAB composites are still optimal in terms of their ability to predict FPPS in each rating. It is possible that some composite other than the current one better predicts FPPS. The definition of FPPS is broad enough to include any number of school pipeline segments, in addition to the "A" School. No research has explored the number of schools in the various pipelines, the nature of the courses and curricula associated with the segments, or whether ASVAB aptitude measures are even relevant in terms of their ability to predict success in segments that have not been included in previous ASVAB validation studies.

Another potential problem is that many pipelines demonstrate extreme differences between the proportions of successes and failures in the sample (e.g., 99% FPP success and 1% FPP failures). Such an extreme split may adversely affect the estimation of the conditional probability of FPPS, particularly in the presence of outliers in the failure sub sample.

A full explanation of why an extreme split may cause problems is beyond the scope of this paper. However, a very brief explanation is as follows: FPPS, when compared to performance measures like FSG that have a continuous, bell-shaped distribution, has a shortcoming when examined from a mathematical and statistical standpoint. The dichotomization of a continuous performance measure necessitates the introduction of an additional nuisance parameter into the analysis, namely the location of the point on the distribution designating the boundary between the successes and failures. When this point is near either extreme of the distribution, then the variance of its estimate is increased, which in turn adversely affects the variances of the slope and intercept parameters in the conditional probability of success estimator (Hannan & Tate, 1965; Prince & Tate, 1966).

Stage II Model

Although the Stage II model is considerably simpler than the Stage I model, it wastefully discards all data except the cut score and PDR bins. In addition, the imposition of linear relationships may introduce bias to the estimation of the conditional probability of FPPS at all points of the distribution, except at the Cut Score and the PDR. The Stage I model, like any estimator, contains estimation error. However, each FPPS rate estimate used to build the Stage I model is unbiased because the properties of the binomial distribution ensure it. The greater the degree of non-linearity demonstrated by the Stage I model, the greater the bias introduced as a result of imposing the Stage II model on top of it. Consequently, the Stage II model is contaminated by both estimation error (inherited from the Stage I model) and bias (from imposing linear relationships that may not have existed in Stage I).

Stage III Model

Adjusting the Stage II model for difficulty in satisfying the FPPS criterion is a reasonable standardization technique. However, a broader, more stable school difficulty measure than FPPS rate in the PDR bin should be used. The overall FPPS rate in the school sample seems more reasonable.

Bin Model Evaluation

Both subjective and objective evaluations of the Bin models were performed. Subjective evaluations were performed by a committee consisting of Janet Held and Geoff Fedak of Navy Personnel Research, Studies, and Technology (NPRST), and the author. Each committee member studied graphical displays of the 70 bin models and judged whether each display indicated the presence or absence of a PDR. A majority vote on each display indicated that a PDR was present in 25 out of the 70 models (35.7%).

The objective evaluation consisted of a statistical comparison of each bin model with a model developed using a baseline methodology. In the author's opinion, an objective evaluation of the bin construction process required a baseline methodology for estimating the conditional probability of FPPS at a given composite score. The bin and the baseline methodologies were compared on the basis of the accuracy of their respective predictions of conditional probability of FPPS. Two logistic regression model prototypes, the quadratic logistic regression model (QLRM) and the linear logistic regression model (LLRM), were selected for the baseline role. Logistic regression is a standard methodology for estimating the conditional mean of a dichotomous criterion variable such as FPPS (Hosmer & Lemeshow, 1989).

The testing procedure described in this section was used to (a) compare the LLRM and QLRM and select which model best describes the relationship between composite score and FPPS in each "A" School sample, (b) determine whether a PDR exists in each sample, and (c) compare the selected LRM (either QLRM or LLRM) with the Bin model and determine whether the Logistic Regression Model (LRM) or Bin model best fits the data.

Inclusion of the QLRM in this study stems from the central role of the PDR concept in RIDE and the need to objectively test for the presence of a PDR. Thus, the choice between LLRM and QLRM provides an objective test for determining whether a PDR exists. If the test indicates that a QLRM (that also has certain characteristics described below) best models the relationship between composite score and FPPS, then there is statistical evidence that a PDR exists. On the other hand, if the test indicates that the LLRM best models the relationship between composite score and FPPS, then there is statistical evidence that a PDR does not exist.

$$S_{i,j}^{Q} = \left\{1 + \exp\left[-\left(\alpha_{2,j} X_{i,j}^{2} + \alpha_{1,j} X_{i,j} + \alpha_{0,j}\right)\right]\right\}^{-1}$$

is the QLRM for the conditional probability of FPPS with respect to individual i in job option j. $\alpha_{k,j}$ is the coefficient of $X_{i,j}^{k}$ (k = 0, 1, 2), and $X_{i,j}$ is the score of individual i on the ASVAB composite for job option j.

The QLRM has exactly one extreme value point, which may be either a maximum or a minimum. As demonstrated in Appendix B, the extreme value point is a minimum if $\alpha_{2,j} > 0$ and is a maximum if $\alpha_{2,j} < 0$. Two distinct QLRM sub models resulted from fitting the generic QLRM to the 70 "A" School samples. One (QLRM #1) is consistent with the assumption of a monotonic increasing relationship between composite score and FPPS over the interval between the cut score (CS) and the maximum observed composite score in the sample (C_{Max}). Hereafter, denote this interval as (CS, C_{Max}]. The second (QLRM #2) is an acceptable QLRM because it is consistent with the PDR concept. Hypothetical curves for QLRM #1 and QLRM #2 are illustrated in Figure 2. It is assumed that CS = 120 and C_{Max}= 160 for all curves in Figure 2.

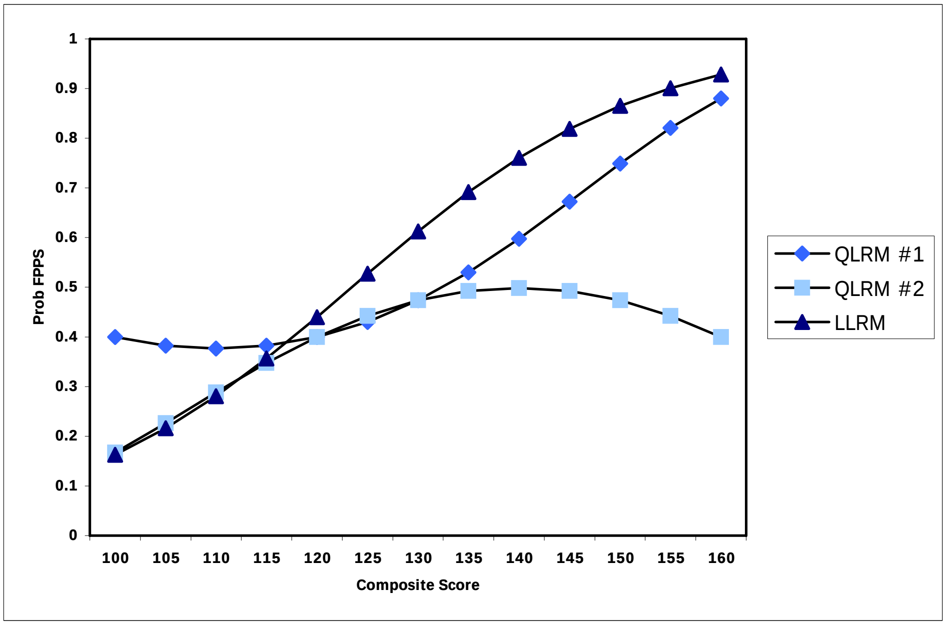

Figure 2. Logistic regression models.

QLRM #1 characteristics: $\alpha_{2,j} > 0$ and so the extreme value point X_{Min} is a minimum. Also, $X_{Max} <$ CS $< C_{Max}$. This is shown in Figure 2, where X_{Min} =110, and so the model is monotonic increasing on [CS, C_{Max}].

QLRM #2 characteristics: $\alpha_{2,j} < 0$ and so the extreme value point X_{Max} is a maximum. In addition, CS $< X_{Max} < C_{Max}$. This is illustrated in Figure 2, where X_{Max} = 140 is a PDR, since the relationship between X and FPPS is monotonic increasing on [CS, X_{Max}] and monotonic decreasing (MD) on[X_{Max}, C_{Max}].

The monotonic character of the LLRM over the entire composite score range makes it appropriate in the context of using aptitude test scores to predict a dichotomous training school success measure such as FPPS.

$S_{i,j}^{L} = \{1 + \exp[-(\beta_{1,j} X_{i,j} + \beta_{0,j})]\}^{-1}$ is the LLRM for the conditional probability of FPPS with respect to individual i in job option j. $\beta_{k,j}$ is the coefficient of $X_{i,j}^{k}$ (k = 0, 1), and $X_{i,j}$ is the score of individual i on the ASVAB composite for job option j. As shown in Figure 2, there is no extreme value point associated with the LLRM, and hence it is monotonic over the entire composite score range. The LLRM is monotonic increasing (monotonic decreasing) if $\beta_{1,j}$ is positive (negative).

The following criteria were used to choose between LLRM and QLRM:

- With the exception of QLRM #2, the model should be monotonic increasing on CS, C_{Max}. This consideration is based upon the assumption that the relationship between composite score and FPPS should, in general, be monotonic increasing.

- P-value test: We choose between LLRM and QLRM based primarily on the p-values associated the highest degree parameter in the respective models. The highest degree parameter of the LLRM is $\beta_{1,j}$, whereas the highest degree parameter of the QLRM is $\alpha_{2,j}$. Comparison of their p-values indicates which parameter we may conclude, with the greatest degree of confidence, is unequal to zero, and, consequently, whether the QLRM or LLRM best fits the data. See Appendix A for discussion on the interpretation of p-values.

Define $\beta_{1,j}$ as the LLRM estimate of $\beta_{1,j}$ and $\hat{\alpha}_{2,j}$ as the QLRM estimate of $\alpha_{2,j}$. For the "A" School sample associated with job option j, we use the p-value to make a preliminary choice between the QLRM and LLRM by applying the following decision rules:

If p-value ($\hat{\alpha}_{2,j}$) < p-value ($\beta_{1,j}$), we may conclude with greater confidence that $\alpha_{2,j}$ is unequal to zero than we could that $\beta_{1,j}$ is unequal to zero. Thus, our preliminary choice is QLRM, which we finalize by performing steps 1 through 3:

1. If the QLRM satisfies the characteristics of QLRM categories #1 or #2 and if no errors were detected during model fit, the model is declared as QLRM. If, in addition, the QLRM satisfies the characteristics of QLRM #2, then a PDR is declared to exist.
2. If (1) is not satisfied, the final model choice is LLRM, provided that $\hat{\beta}_{1,j}$ > 0 and no error conditions were detected during parameter estimation.
3. If (2) is not satisfied, then the model is declared "No Decision," indicating that neither QLRM nor LLRM provides a satisfactory fit.

If p-value ($\hat{\alpha}_{2,j}$) > p-value ($\beta_{1,j}$), then our preliminary model choice is LLRM. That decision becomes final if $\beta_{1,j}$ > 0 and no error conditions were detected during parameter estimation. However, if $\beta_{1,j}$ < 0 or at least one error is detected, the model is declared "No Decision."

Once the QLRM vs. LLRM winner is selected, it is compared with the Stage II bin model. The (non-standardized) Stage II model, rather than the (standardized) Stage III model, is compared with the QLRM-LLRM winner because the basis for comparison is accuracy of conditional probability of FPPS prediction.

The "expected absolute total error" (EATE) criterion was used to compare the Bin and LRM models. As described under *Criticisms of RIDE SPSU Model*, the Bin model construction process introduces both bias and estimation error into its estimate of conditional probability of FPPS. In contrast, the asymptotic unbiased property of maximum likelihood estimators (MLE) means that the logistic regression parameter estimates are unbiased in the limit as the sample size becomes large (Stuart & Ord, 1991). (The author is not aware of any studies indicating whether, for a fixed sample size, the LRM parameter MLEs are still unbiased and, if not, the degree of bias present.)

Preliminary Bin vs. LRM comparisons were performed using 95 percent confidence intervals. Overall, these results indicated that the LRM had slightly narrower confidence interval widths than the Bin model. However, since estimator bias is not considered in the confidence interval calculation, a criterion was sought that would incorporate both bias and estimation error variance into the comparison. The EATE criterion was developed by assuming that ε (epsilon, the total FPPS rate estimation error due to the presence of both bias and FPPS rate estimation error variance) is normally distributed with mean equal to the bias and variance equal to the estimation error variance. Mathematically, EATE is the expected value of the absolute value of ε ($E|\varepsilon|$) and is given by

$$\text{EATE} = E|\varepsilon| = 2\sigma\varphi\left(-\frac{\mu}{\sigma}\right) - 2\mu\Phi\left(-\frac{\mu}{\sigma}\right) + \mu \text{, where}$$

σ^2 is estimation error variance,

μ is the bias of the estimator,

$\varphi(\,)$ is the standard normal probability density function, and

$\Phi(\,)$ is the standard normal cumulative distribution function.

This formula is used to calculate EATE of the logit in the LRM model, and the EATE of the Bin model. The derivation of EATE is given in Appendix A, as are the procedural details of the Bin vs. LRM comparison.

Table 1 summarizes the results of the analyses. The Bin columns indicate the mean of the Bin model EATEs for each rating. Each mean was computed by averaging the EATEs over all integer composite scores between the cut score and the C_{Max} in that rating. The LRM columns indicate the mean of the LRM model EATEs for each rating, again computed by averaging over all composite scores between the cut score and C_{Max} for that rating. When these columns were averaged over all ratings, the mean Bin EATE was 0.050 and the mean LRM EATE was 0.030. The B/L columns indicate whether the Bin model or LRM model produced the smaller EATE. In this comparison, the Bin model produced the smaller EATE only 7 times, while the LRM produced the smaller EATE 62 times. The Model column indicates which LRM (QLRM or LLRM) was the superior LRM for that rating and was matched against the Bin model in the EATE comparison. The appearance of (PDR) in that column indicates that the chosen QLRM satisfied the criteria for the existence of a PDR.[1] Six of the 70 LRMs were QLRM, 55 of them were LLRM, and the remaining 9 were "No Decision."[2] Five of the six ratings that satisfied the QLRM criteria also satisfied the conditions for the existence of a PDR. These ratings are designated by an asterisk (*) in the Model columns.

[1] Note: The LLRM was matched against the Bin model whenever the QLRM vs. LLRM comparison resulted in a "No Decision" outcome.

[2] No Bin vs. LRM comparison was performed for the PH 5Y rating because only one bin resulted when the bin construction procedure was performed on that sample. At least 2 bins are necessary to calculate the Stage I estimate.

Table 1
Bins vs. LRM

Rate	Bin	LRM	B/L	Model	Rate	Bin	LRM	B/L	Model
AB--GE	.065	.048	L	QLRM*	AC--5Y	.097	.037	L	LLRM
AD--SG	.012	.010	L	LLRM	AE--SG	.022	.013	L	LLRM
AECFAE	.048	.015	L	LLRM	AG--SG	.100	.032	L	LLRM
AK--SG	.027	.021	L	NoDe	AM--GE	.013	.008	L	LLRM
AO--SG	.061	.020	L	LLRM	AS--SG	.119	.056	L	NoDe
AT--GE	.042	.017	L	LLRM	AZ--SG	.020	.013	L	LLRM
BU--5Y	.085	.042	L	QLRM*	CE--5Y	.040	.020	L	LLRM
CM--5Y	.064	.025	L	LLRM	CTA-SG	.017	.017	L	LLRM
CTI-SG	.119	.031	L	LLRM	CTM-AE	.046	.024	L	LLRM
CTO-SG	.032	.033	B	QLRM*	CTR-SG	.053	.035	L	NoDe
CTT-SG	.015	.010	L	LLRM	DC--SG	.023	.013	L	LLRM
DK--SG	.060	.035	L	NoDe	DT--GE	.021	.010	L	LLRM
EA--5Y	.055	.039	L	LLRM	EM--SG	.045	.022	L	LLRM
EN--SG	.060	.023	L	LLRM	EN--AT	.108	.035	L	LLRM
EO--5Y	.015	.010	L	LLRM	ETS-GE	.051	.019	L	LLRM
EW--SG	.036	.027	L	LLRM	EW--AE	.138	.048	L	LLRM
FT--GE	.028	.024	L	LLRM	GM--SG	.053	.034	L	LLRM
GSE-GE	.083	.040	L	LLRM	GSM-GE	.075	.028	L	LLRM
HM--GE	.027	.008	L	LLRM	HT--GE	.028	.023	L	QLRM
IC--GE	.111	.032	L	LLRM	IS--SG	.043	.025	L	LLRM
JO--5Y	.058	.044	L	LLRM	LI--SG	.025	.050	B	LLRM
MM--SG	.029	.014	L	LLRM	MM--NF	.069	.085	B	LLRM
MMS-SG	.078	.020	L	LLRM	MN--SG	.038	.029	L	NoDe
MR--SG	.040	.036	L	LLRM	MS--SG	.072	.019	L	NoDe
MSS-SG	.070	.049	L	LLRM	MT--AE	.036	.027	L	LLRM
----NF	.016	.022	B	LLRM	OS--SG	.008	.007	L	LLRM
PH--5Y	----	----	---	QLRM*	PN--SG	.027	.020	L	LLRM
PR--SG	.011	.015	B	LLRM	QM--SG	.044	.038	L	LLRM
RM--SG	.029	.012	L	LLRM	RP--SG	.048	.048	B	NoDe
SH--SG	.032	.021	L	LLRM	SK--SG	.028	.019	L	NoDe
SKS-SG	.101	.094	L	LLRM	SM--SG	.030	.024	L	LLRM
SS--SF	.040	.016	L	LLRM	STG-GE	.007	.007	L	LLRM
STS-GE	.068	.022	L	LLRM	SW--5Y	.074	.042	L	NoDe
TM--SG	.065	.041	L	LLRM	UT--5Y	.037	.038	B	QLRM*
YN--SG	.033	.013	L	LLRM	YNS-SG	.094	.071	L	LLRM

School Success Component (SSC) of CLASP

The school success utility component predicts "A" School success as a function of the operational ASVAB selector composite for each rating. Prior to CLASP, classifiers made "A" School assignments based on the cut score for each rating, without considering the degree to which the applicant may exceed that score. Based upon the assumption that an applicant's likelihood of success increases with aptitude test score, the school success component was designed to incorporate information about the complete range of scores, instead of focusing solely on whether the cut score was satisfied. For the original CLASP implementation in the early 1980s, Navy validation samples were obtained from Paul Foley of Navy Personnel Research and Development Center (NPRDC). Linear regression analyses were performed to develop unique school success equations for ratings in which validation data was available. Thus, in the original CLASP implementation, different selector composites were used to predict school success for different ratings. The original equations were typically characterized by non-integer weights and, in some instances, negative weights.

In 1984, a new policy allowed only operational ASVAB selector composites to be used as school success equations in CLASP. Therefore, the current school success equation for each job option is identical to the ASVAB composite currently used for selection purposes. Accordingly, CLASP school success criterion measures vary from job option to job option. For a given job option, the school success criterion is determined by the "A" School performance measure used in the ASVAB validation study that recommended use of that particular composite. Whenever an ASVAB validation study recommends that the ASVAB composite(s) currently used for selection and/or the associated cut score(s) be replaced, NPRST immediately submits for operational CLASP implementation an updated school success mean and standard deviation for each job option associated with the rating. When an ASVAB selector composite change is recommended and approved, Commander, Navy Recruiting Command (CNRC) then changes the corresponding school success equation(s) in the operational CLASP implementation. Several "A" Schools select students using multiple composites and cut scores, either as a multiple hurdle or as an "either/or" criterion. For CLASP job options associated with these ratings, one composite is designated by NPRST as the CLASP school success equation.

The standardized school success payoff for individual i in rating j is given by

$$S_{i,j}^{*} = 50 + 10\left(\frac{S_{i,j} - \mu_{SS,j}}{\sigma_{SS,j}}\right) \qquad \text{Sch\_Suc}$$

where

$S_{i,j}^{*}$ is the standardized school success payoff associated with placing individual i in rating j,

$S_{i,j}$ is the ASVAB composite score for individual i in rating j,

$\mu_{SS,j}$ is the $S_{i,j}$ reference population mean for rating j, and

$\sigma_{SS,j}$ is the $S_{i,j}$ reference population standard deviation for rating j.

The reference population used to estimate the mean and standard deviation consists only of recruits who satisfy the ASVAB selector criteria for that rating, not the entire recruit population. Subtracting $\mu_{SS,j}$ from $S_{i,j}$ and dividing that difference by $\sigma_{SS,j}$ in equation Sch_Suc adjusts each $S_{i,j}$ for differences across ratings in the average ability level required to qualify for and successfully complete "A" School. As a result, equation Sch_Suc transforms the $S_{i,j}$ into a common metric for all ratings and facilitates comparison across ratings for individual i. However, it is not known if conversion to this common metric is sufficient to completely eliminate the tendency for the easier schools to experience higher school success utility scores, on the average.

CLASP Parameter Update Considerations

The reference population means and standard deviations for each job category are the only School Success component parameters subject to updating. The CLASP parameter update software automatically generates an update for these parameters during the annual CLASP parameter update. In addition, NPRST possesses a software package to update any specified subset of the school success mean and standard deviation parameters when ASVAB selector composite and/or cut score changes have been recommended and approved.

RIDE SPSU and CLASP SSC Summary

The section closes with a discussion of several important considerations in building and maintaining the SPSU component of RIDE. Also included is a description of strengths and weakness of SSC and SPSU.

CLASP SSC Weaknesses

School success equations in the current CLASP implementation are chosen from a short list of approximately 12 ASVAB (unique) selector composites. This small number of unique composites, relative to the approximately 120–130 job options currently sold in CLASP, means that the same composite is used for several job options. For example, as of March 2003, CLASP used Verbal and Arithmetic Reasoning (VE+AR) to predict school success in 17 job options and Arithmetic Reasoning, Math Knowledge, Electronics Information, and General Science (AR+MK+EI+GS) in 30 job options. Hence, the SSC has a limited differential prediction capability, meaning that it cannot distinguish differences in school success utility between pairs of job options using the same equation. A partial solution may be achieved in job options that use multiple composites for selection, either as a multiple hurdle or as an "either/or" criterion. If appropriate weights could be found, additional school success equations could be created by taking a weighted sum of all composites appearing in the "A" School selection standards for these job options. The number of CLASP job options sharing the same composite could be reduced substantially. In addition to concerns regarding the quality of differential prediction, the SSC lacks the flexibility to implement anything other than a linear relationship between composite score and utility.

CLASP SSC Strengths

The advantage of the SSC is that only Navy applicant data from PRIDE is required to update the CLASP parameters, including SSC mean and standard deviation parameters. When an ASVAB validation study recommends a selector composite and/or cut score change for a given rating, CLASP does not require a new prediction model. CLASP requires only that mean and standard deviation parameters for that rating be updated based upon the new composite and/or cut score. NPRST uses a simple procedure to estimate the new parameters and forward them for implementation. "A" School validation samples are not required for this purpose.

RIDE SPSU Weaknesses

Development and maintenance of bin and/or logistic regression models for predicting FPPS requires an "A" School validation sample for each rating. As is currently the case with CLASP, when an ASVAB validation study recommends a change to the operational selector composite in a given rating, a corresponding change to the SPSU component of RIDE will be required. However, unlike CLASP, the RIDE parameter update requires estimation of both a new PDR and the FPPS rate at the new PDR. School performance data would be required to accomplish this. In addition, it is anticipated that in some situations, such changes may be more difficult and time-consuming than is currently the case with CLASP. When a selector composite change is recommended and approved for a given rating, it may not be advisable to immediately develop and implement a new FPPS prediction model for that rating using the currently available validation sample and the replacement (i.e., new) selector composite. This will be especially true if the incumbent and replacement composites will select student populations that are significantly different from one another. Accordingly, it may not be feasible to implement the new selector composite in RIDE until after sufficient students have been selected with the replacement composite to develop and implement a new prediction model.

RIDE SPSU Strengths

Availability of "A" School performance data will facilitate development of non-linear models of the relationship between composite score and school performance. It will also facilitate development of unique SPSU equations for more job options than is currently feasible in CLASP. Although previous sections raised several questions concerning the quality of the FPPS criterion and the quality of the Bin and LRM estimators of the conditional probability of FPPS, the availability of school performance data would facilitate further research on criterion measure alternatives to FPPS.

CLASP Aptitude/Difficulty and RIDE AFQT Component Comparison

In ascertaining whether an applicant is suited to a particular job, the employer must assess the job's requirements and the applicant's abilities. The employer must decide whether the prospective employee has the abilities required to succeed in the job.

During a typical employment interview, the employer judges the applicant's abilities using some internal scale. The internal scale may not be well defined, but allows the employer to evaluate and rank-order prospective employees. The employer can be more certain about the characteristics of the job and the type of person most likely to fill the job successfully. The employer's experience enables him to rank-order jobs based on the technical ability they require. This continuum forms a second scale. For example, an employer may judge that a particular applicant belongs to the upper 25 percent of applicants, as assessed on the internal aptitude scale. A particular job may be rated by the employer as belonging to the upper 25 percent of jobs on the scale of technical aptitude required to succeed. Having established the relative positions of both the job and the applicant on their respective scales, the employer may judge their correspondence to each other. In this case, there appears to be a match and the applicant will likely be offered the job.

The Aptitude/Difficulty component of CLASP works similarly to the employer's evaluative techniques. This utility function involves two scales: (1) a measure of an applicant's overall technical aptitude, and (2) a measure of the rating's technical difficulty or complexity. Thus, given an applicant's technical aptitude and a rating's technical difficulty, the utility of that person-job match may be evaluated and compared with other possible person-job matchups.

Kroeker and Rafacz (1983) provide details concerning the technical aptitude and job difficulty scales. The technical aptitude composite (TAC), computed as MC+AS+EI+GS, measures the applicant's technical aptitude for purposes of the Aptitude/Difficulty component. The following equation transforms the TAC so the resulting transformed aptitude score (TAS) is between 40 and 100, inclusive.

$$A_i = 40 + 60\left(\frac{C_i - 180}{280 - 180}\right) \qquad \text{Apt\_Dif\_TAS}$$

Truncate to A_i = 100 if $C_i \geq 280$ and truncate to A_i = 40 if $C_i \geq 180$, where A_i and C_i are the TAS and TAC scores for individual i, respectively. The TAS distribution must fall in this range because the Aptitude/Difficulty utility function described below is constructed such that its aptitude argument must satisfy this property.

As indicated by equation Apt_Dif_TAS, the transformation truncates TAC scores that are either less than 180 or greater than 280 so that they fall at the extremes of the TAS distribution. Since the minimum standard score of each subtest is 20 and the maximum standard score is 80, the minimum TAC score is 4 x 20 = 80 and the maximum is 4 x 80 = 320. Consequently, TAC scores between 80 and 180 correspond to a score of 40 on the TAS, while TAC scores between 280 and 320 correspond to a TAS score of 100. The original rationale for the truncation in equation Apt_Dif_TAS is unknown, but its apparent effect is to prevent the TAS distribution from being tightly concentrated around its mean and to spread it more uniformly over the range between 40 and 100.

The job difficulty scale was established using paired comparison methodology. (Kroeker, personal communication, 1998). Initial scale values were produced for the complete job set by applying the paired comparison procedure to two data sets: (1) experimenter judgments about the cognitive skills required by each job, and (2) experimenter estimates of the visual perceptual attributes required. Data were then collected from subject matter experts (SMEs) who were asked to compare the job difficulty of small groups of ratings. The SMEs ranked the difficulty of 8 to 10 jobs in pairs, thus contributing to a matrix from which new scale values could be derived for the entire job set. The scale was then modified by using an iterative procedure to revise psychological values (Kroeker, 1982).

The unstandardized technical aptitude/job difficulty utility associated with assigning person *i* to job *j* is given by:

Equation Apt_Dif_Util:

$$U_{A/D}(A_i, D_j) = B_{0,0} + B_{2,0}(A_i - 100)^2 + B_{0,1}(D_j - 35) + B_{2,2}(A_i - 100)^2(D_{j-35})^2 + B_{2,1}(A_i - 100)^2(D_j - 35) + B_{0,2}(D_j - 35)^2 \qquad \text{where}$$

$B_{0,0} = 30.0$, $B_{2,0} = -0.0005$, $B_{0,1} = 1.867$, $B_{2,2} = -0.00001696$, $B_{2,1} = -0.0001867$ and $B_{0,2} = -0.01244$, $U_{A/D}(A_2, D_j)$ is the raw Aptitude/Difficulty utility of assigning person *i* to job *j*, A_i is the TAS score of person *i*, and D_j is the job difficulty of rating *j*.

The following briefly explains the development of equation Apt_Dif_Util. Ward (1977) is an excellent source reference for this topic. The classification policymaker assumed that $U_{A/D}(A_2, D_j)$ is a polynomial in two variables: applicant aptitude A_i and job difficulty D_j. The maximum degrees of A_i and D_j of the (bivariate) polynomial are determined by the number of initial conditions, as described below. Hence, it was originally specified as

$$U_{A/D}(A_i, D_j) = \sum_{i=0}^{2}\sum_{j=0}^{2} B_{i,j}(A_i - 100)^i (D_j - 35)^j \qquad \text{Apt\_Dif\_Poly}$$

There are (2+1) x (2+1) = 9 unknown coefficients ($B_{0,0}$, $B_{0,1}$, $B_{0,2}$, $B_{1,0}$, $B_{1,1}$, $B_{1,2}$, $B_{2,0}$, $B_{2,1}$, and $B_{2,2}$) to be determined. Step 2 specifies a set of initial conditions (either on the utility function itself or on its partial derivatives) at critical values of A and D. For example, equation (Apt_Dif_Poly) was developed using a set of initial conditions similar to the following:

(1) $U_{A/D}(40,40) = 32.34$

(2) $\frac{\partial U_{A/D}}{\partial D}$ = 0 when evaluated at A = 40, D = 43.1

(3) $U_{A/D}(40,100) = -204.65$

(4) $U_{A/D}(100,100) = 98.8$

(5) $U_{A/D}(100,40) = 39.02$

(6) $\frac{\partial U_{A/D}}{\partial D}$ = 0 when evaluated at A = D = 100

(7) $U_{A/D}(70,40) = 37.35$

(8) $U_{A/D}(70,100) = 22.93$

(9) $\frac{\partial U_{A/D}}{\partial D}$ = 0 when evaluated at A = 70, D = 65.7

Note the number of initial conditions equals the number of unknown coefficients. When the initial conditions are substituted into Apt_Dif_Poly, the result is a system of 9 linear equations in the 9 unknown coefficients. The coefficients may be determined by solving the linear system.

As described by Ward (1977), the initial conditions are based upon policymaker requirements regarding the desired behavior of the function at pre-specified values of *A* and *D*. Judicious choices in the initial condition specification will give $U_{A/D}(A,D)$ its desired appearance over the entire range of allowable values of A and D.

As far as the author can determine, the Aptitude/Difficulty, Priority/Preference, and Attrition Component Utility functions were all developed as mathematical representations of personnel classification policy. For example, the Aptitude/Difficulty utility function is based upon policymaker judgments concerning the value to the Navy of assigning an individual with a given technical aptitude level to a job with a given level of technical difficulty. The A/D, P/P, and Attrition utility functions do not appear to either represent the outcome or results of any empirical study or to be motivated by any such study. The author is not aware of any research, either inside or outside the military, which has produced an empirically-based model describing utility as a function of a person attribute and a job attribute. As described below, similar procedures were used to determine the coefficients for the raw utility functions in the Priority/Preference and the Attrition components of CLASP.

Figure 3 shows a graph of equation Apt_Dif_Util with $U_{A/D}(A,D)$ plotted as a function of Job Difficulty for fixed applicant Aptitude values A = 40, 50, 60, 80, 90, and 99. The uppermost curve on Figure 3 represents the utility values for the highest technical aptitude level (99) across the entire range of job difficulty. The region at which the curve assumes its maximum value occurs at the upper end of the difficulty scale. This implies the utility function tends to assign the highest aptitude individuals to the most technically complex ratings. The curve's gradual downward slope from the region of greatest technical difficulty to the region of least technical difficulty implies that smaller utility values are awarded when high-aptitude individuals are assigned to low-difficulty jobs. Although the probability of such assignments is reduced accordingly, they may still take place, due to the influence of the other CLASP components. The lowest curve represents the utility values for the lowest technical aptitude level (40) across the entire range of job difficulty. Its maximum value occurs at the lowest end of the difficulty scale; its sharply downward slope in the direction of increasing job difficulty means that low-ability applicants will almost always be assigned to the least complex jobs.

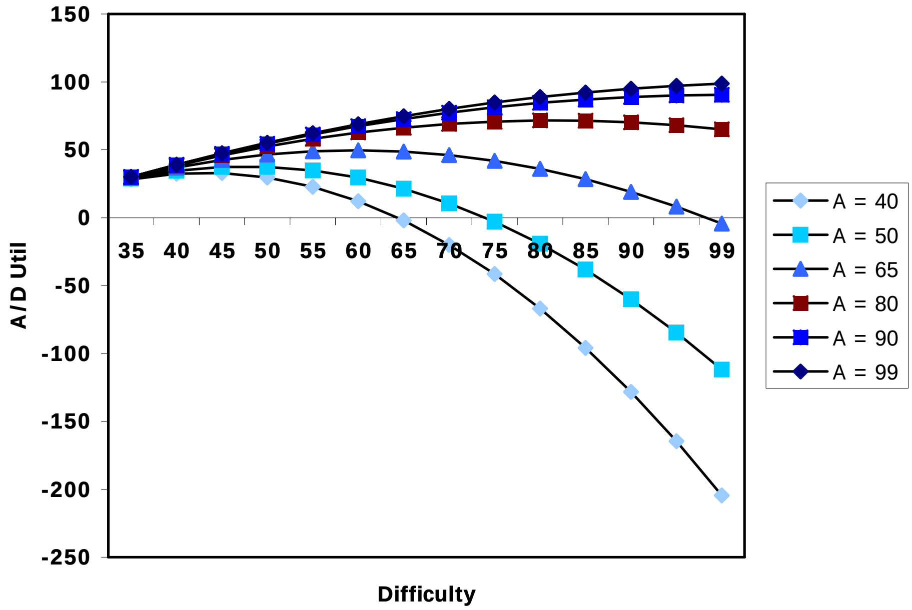

Figure 3. Aptitude/Difficulty Utility.

The middle curve indicates that applicants of average ability (65) have a reasonable chance to be assigned to ratings of all difficulty levels. However, given that the maximum of this curve occurs in the range of intermediate job difficulty, it is most likely they will be assigned to ratings of intermediate technical difficulty.

Thus, for a given level of applicant aptitude, the Aptitude/Difficulty component awards the largest utility values to assignments providing the closest correspondence between the applicant's ranking on the technical aptitude score distribution and the job's ranking on the job difficulty distribution. In other words, the largest utility values are awarded when high aptitude applicants are matched up with most difficult jobs. Intermediate aptitude applicants are awarded the largest utility when they are matched with intermediate difficulty jobs, although the utility of this matchup is not as large as that of the high aptitude individual and high difficulty job. Low aptitude applicants are awarded the largest utility when they are matched with low difficulty jobs, although the utility of this matchup is not as large as that of the intermediate aptitude individual and intermediate difficulty job. Table D-1 in Appendix D shows, for each fixed A, the difficulty level $D_{Max}(A)$ that maximizes $U_{A/D}(A,D)$. That is, for any fixed A, $D_{Max}(A)$ is the difficulty level D such that $U_{A/D}(A, D_{Max}(A)) > U_{A/D}(A,D)$ for all D in [40,99]. As shown therein, both $D_{Max}(A)$ and $U_{A/D}(A, D_{Max}(A))$ are increasing functions of A.

In Figure 3, the 6 curves for the 6 aptitude levels do not intersect. Thus, for a given job difficulty level, the utility associated with assigning an applicant with aptitude A_1 is greater than the utility associated with TAS score A_2 if $A_1 > A_2$. Stated differently, larger applicant aptitude levels result in larger utility values, regardless of job difficulty level. As a general rule, this seems reasonable, with the possible exception of the lowest job difficulties. One may argue that higher applicant aptitudes should result in smaller utilities for the lowest job difficulties, since the assignment of high aptitude individuals to these jobs wastes talent that could productively be used in the technically more difficult jobs. Such an argument could be used in support of the AFQT component in RIDE.

The standardized Aptitude/Difficulty payoff is calculated as:

$$U^*_{A/D}(A,D) = 50 + 10\left(\frac{U_{A/D}(A,D) - \mu_{AD}}{\sigma_{AD}}\right) \qquad \text{Apt\_Dif\_Std}$$

CLASP Parameter Update Considerations

μ_{AD}, σ_{AD}, and the job difficulty (i.e., job complexity) index parameters, D_j, for each rating constitute the Aptitude/Difficulty component parameters subject to updating. The CLASP parameter update software automatically generates updates for μ_{AD} and σ_{AD} during the annual CLASP parameter update. Kroeker (personal communication, 1998) documents the procedures and methodology he used to update the original set of job difficulty parameters he developed in the late 1970s or early 1980s. The author knows of no reason why these updated parameters could not be implemented in CLASP at this time.

RIDE AFQT Utility

The purpose of the RIDE AFQT Component is to "penalize" the applicant's utility scores in ratings where the AFQT score suggests the applicant is over-qualified. If the degree of over-qualification is large enough, both the Navy's and the applicant's interests are best served by placement in a rating in which the applicant’s general aptitude more closely matches that of other applicants assigned to the rating. This concept is based on the assumption that the AFQT score represents a measure of the applicant's overall, general aptitude, while the ASVAB selector composite score measures specific skills and aptitudes for the rating.

Figure 4 demonstrates this concept. The maximum AFQT utility is achieved by individuals whose AFQT score is ≤ the mean AFQT score M_j of individuals assigned to the rating. Utility decreases from a maximum of Q_{Max} = 100 to a minimum of Q_{Min} = 0 as the individual's AFQT score substantially exceeds the mean AFQT score for that rating.

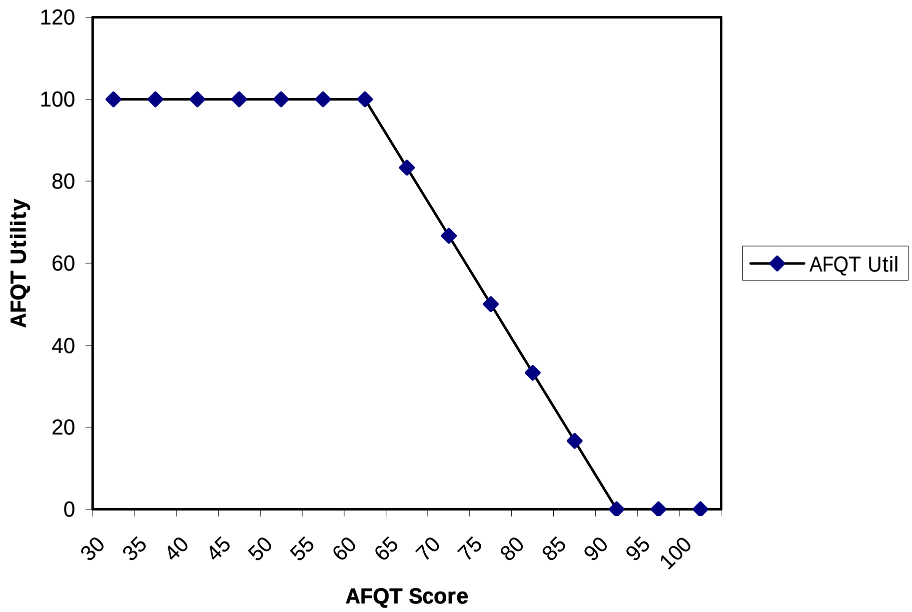

Figure 4. RIDE AFQT Utility.

Define $Q_{i,j}$ as the AFQT utility associated with assigning individual i to job option j. Define A_i as the AFQT score of individual i, M_j as the mean of the AFQT distribution in job option j, and σ_j as the standard deviation of the AFQT distribution in job option j. In addition, define $\sigma_j \geq 0$ as the offset from M_j that defines the maximum AFQT score for which $Q_{i,j} = Q_{Max} = 100$. Also, define $\theta_j > M_j + \delta_j$ as the minimum AFQT score for which $Q_{i,j} = Q_{Min} = 0$. In other words, as A_i increases from 0 to 100, $M_j + \delta_j$ represents the AFQT score at which the penalty begins to take effect, while θ_j is the AFQT score at which the penalty first reaches its maximum.

In Figure 4, $M_j = 55, \quad \sigma_j = 10, \delta_j = \frac{\sigma_j}{2}$, and $\theta_j = M_j . + 3.5\sigma_j$.

Then $M_j + \delta_j = M_j + \frac{\sigma_j}{2} = 60$ and $\theta_j = M_j + 3.5\sigma_j = 90$. Note that the specifications for $M_j + \delta_j$ and θ_j have been modified since the Folchi (1999) specification. $Q_{i,j}$ may then be defined as follows:

$$Q_{i,j} = Q_{Max} \quad \text{if} \quad A_i \leq M_j + \delta_j,$$
$$Q_{i,j} = Q_{Min} \quad \text{if} \quad A_i \geq \theta_j, \text{ and}$$
$$Q_{i,j} = \frac{Q_{Max} - Q_{Min}}{M_j + \delta_j - \theta_j}\left[A_i - \left(M_j + \delta_j\right)\right] + Q_{Max} \qquad \text{if} \qquad M_j + \delta_j \leq A_i \leq \theta_j .$$

RIDE AFQT and CLASP Aptitude/Difficulty Summary

In summary, $U_{A/D}(A,D)$ is a mathematical representation of a classification policy that assigns each applicant to a rating whose technical difficulty, D, approximately corresponds with technical aptitude A. The costs (both to the Navy and the applicant) of a mismatch seem clear. Worker boredom and a lost opportunity to assign individuals to jobs that better match their skills and aptitude are the costs associated with assigning applicants to jobs that are too easy. Decreased productivity is the cost of assigning applicants to jobs for which they lack the required aptitude to perform properly.

Comparison of Figures 3 and 4 indicates the CLASP A/D and RIDE AFQT components are quite different. RIDE AFQT penalizes for "over-qualification" in a given rating (as measured by the degree to which the applicant's AFQT score exceeds the $M + \delta$ point in that rating's AFQT distribution). RIDE imposes no such penalty for under-qualification. In contrast, CLASP A/D penalizes for "under-qualification" (as measured by the degree to which the applicant's technical aptitude measure is less than the Job Difficulty measure in that rating), but imposes no penalty for over-qualification. Regardless of rating, the RIDE AFQT utility function is monotonically decreasing (flat between the minimum AFQT score and $M + \delta$, downward sloping between $M + \delta$ and θ, and then flat between θ and the maximum AFQT score). In contrast, Figure 3 shows the CLASP A/D function is monotonically increasing between the minimum and maximum values of A. As the Job Difficulty increases, $U_{A/D}(A,D)$ increases more rapidly between the minimum and maximum values of A. The CLASP policy that rewards a larger aptitude with a larger utility value, regardless of job difficulty level, is not present in the RIDE AFQT model. RIDE rewards larger aptitudes with larger utility scores only in the more difficult jobs.

In summary, the RIDE AFQT and CLASP A/D components seem motivated in conceptually opposite directions. Two possible methods for judging and comparing them are: (a) evaluate them in context with the remaining model components, and (b) evaluate them from a policymaker's standpoint. One possible technique of accomplishing (a) is to apply the two algorithms to a baseline set of applicant records and, applicant by applicant, compare the RIDE and CLASP optimal lists. In particular, since RIDE requires less applicant input information, this technique could provide useful insights into the manner in which the SPSU and AFQT components interact to generate a composite RIDE utility. It may also be helpful in understanding how the School Success and Aptitude/Difficulty components of CLASP interact, and how the RIDE composite utilities compare with a composite of the School Success and Aptitude/Difficulty components of CLASP. Evaluation of (b) requires a policymaker to express opinions on questions such as: Do the "under-qualification" and "over-qualification" concepts make sense in the Navy environment? In particular, should under-qualified (or over-qualified) applicants be awarded fewer utility points if their technical aptitude does not closely match the technical difficulty level of a given rating, under the premise that too low (or too high) an aptitude level makes them less likely to succeed in that rating?

The 4 remaining CLASP components that do not have counter-parts in RIDE are discussed in the following sections.

Navy Priority/Individual Preference Component

Kroeker and Rafacz (1983) provide an excellent introduction to the Priority/Preference component and description of the priority scale.

> The purpose of this component is to incorporate both Navy priorities and individual preferences when assigning recruit applicants to ratings. These two sets of objectives may be incompatible, particularly if both are described by utility functions allowed to vary independently. For example, the gain in utility resulting from an applicant's expression of strong preference for a particular rating may be offset by a loss in utility if the rating has a low Navy priority.
>
> To overcome the deficiency of a strictly additive model, an interactive utility function was designed. Thus, a utility value is obtained as a function of the Navy priority index for a particular rating in conjunction with the applicant's specified preference value for that rating. To address both Navy priority and individual preference, two scales were derived:
>
> Priority Scale: Navy priorities were obtained from the career reenlistment objectives listed by the Office of the Chief of Naval Operations. These priorities were augmented and modified using rating popularity and rating size as variables in a least squares regression analysis. The resulting priority scale was refined by data collected from 10 Navy personnel managers concerned with setting recruiting goals and "A" School priorities. In a procedure similar to that used to establish the job complexity scale, these officers compared the relative importance to the Navy of small groups of ratings, by pairs. As with the job complexity scale, values were then modified using a procedure to revise estimates of psychological scale values (Kroeker, 1982).

The Kroeker and Rafacz description of the individual preference scale is not consistent with the actual CLASP implementation. Therefore, the following alternative description is provided:

An individual preference value is computed for each rating. The applicant classification process at the Military Entrance Processing Station (MEPS) does not allow enough time for the recruit to rank order all ratings s/he may potentially be assigned to. Therefore, preference values are not determined on the basis of individual ratings. However, since each rating belongs to exactly 1 of approximately 15 occupational group categories, preferences are determined by asking the applicant to rank order up to 5 occupational groups in terms of preference. Each rating in the most preferred occupational group receives the highest possible preference value (100), each rating in the second ranked group receives the second highest possible preference value (90), etc. Thus, the preference scale can be expressed as

$$Z_{i,j[r]} = 100 - 10(r - 1),\ r = 1, 2, \ldots, n_p, \text{ where}$$

$Z_{i,j[r]}$ is the individual preference value for individual i in rating $j[r]$,

the index $j[r]$ ranges over all ratings in the rth ranked occupational category,

r is the occupation group ranking, and

$1 \leq n_p \leq 5$ is the number of occupational groups the applicant expresses a preference for.

For ratings in the remaining occupational groups r for which the applicant did not express a preference, $Z_{i,j[r]}$ is assigned the lowest possible preference value (20). Thus, the individual preference scale ranges between 20 and 100, with larger preference values associated with the applicant's most preferred ratings and smaller values associated with his/her least preferred ratings.

Given the Navy priority index of a rating and the individual's preference value of the rating, the unstandardized Priority/Preference utility is given by Equation Prior_Pref_Unstd:

$$U_{PP}\left(W_j, Z_{i,j}\right) = 90.0 + (0.001)\, W_j^2 + (1.8)\,(Z_{i,j} - 100) - (0.0000014)\, W_j^2\, (Z_{i,j} - 100)^2 - (0.00018)\, W_j^2\, (Z_{i,j} - 100) + (0.009)\,(Z_{i,j} - 100)^2, \text{ where}$$

$U_{PP}\left(W_j, Z_{i,j}\right)$ is the priority/preference utility associated with individual i in rating j,

W_j is the Navy priority index value for rating j, and

$Z_{i,j}$ is the individual preference value for individual i in rating j.

In Figure 5, $U_{pp}(W_j, Z_{i,j})$ is plotted on the vertical axis against Individual Preference on the horizontal axis, for priority values of 100, 80, 50, and 0. The four curves are non-intersecting and appear in order of increasing priority level from bottom to top. Thus, for any fixed individual preference value, a larger priority value generates a larger priority/preference utility than a smaller priority value.

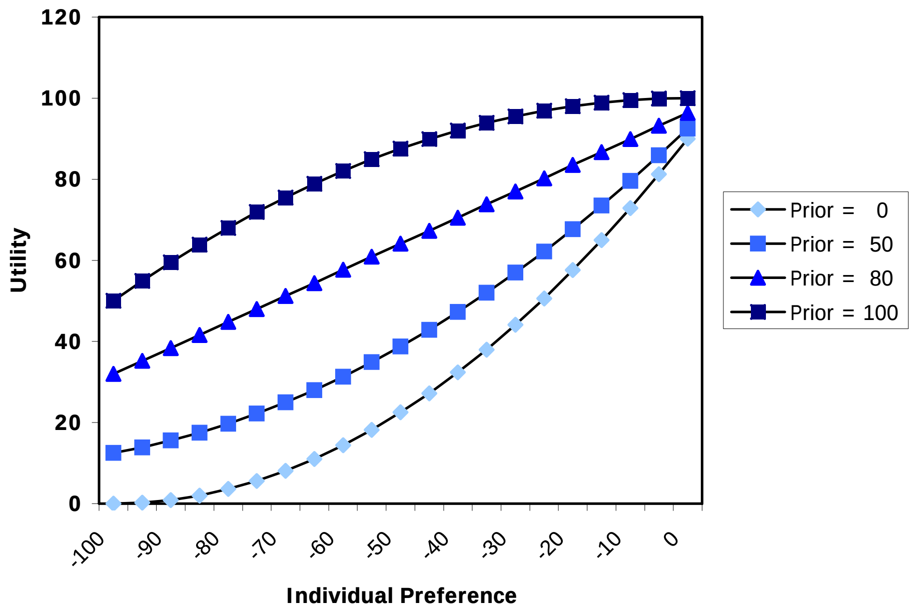

Figure 5. Priority/Preference Utility.

In addition, the utility for each priority level is an increasing function of individual preference level. Thus, the utility of a person-rating match increases both as a function of the rating priority (for a fixed individual preference level) and as a function of the individual's preference for the rating (for a fixed priority level). However, since the curves are not parallel, utility is a non-linear function of priority and preference.

The uppermost curve represents utility values corresponding to the highest level of Navy priority (100) across the entire range of individual preferences. A strong or moderate preference for a high priority rating yields a high utility value, since both the Navy's and the applicant's interests are satisfied by such an assignment. A low preference for a high priority rating yields a moderate level utility that expresses the importance of the rating to the Navy. The lowest curve represents utility values corresponding to the lowest Navy priority level (0) across the range of individual preferences. A strong preference for a low-priority rating produces a high utility because of the Navy's attempt to honor the applicant's preference. A moderate degree of preference for the rating, however, results in a relatively low utility value because the Navy's interests are not served by such an assignment. An expression of no preference for a low-priority rating results in the lowest possible utility level because neither the Navy's nor the applicant's interests are satisfied.

Equation Prior_Pref_Unstd was developed by assuming that $U_{pp}(W, Z)$ is a polynomial in Navy priority W and individual preference Z. It was assumed to be a second degree polynomial in both W and Z, and thus it has 9 unknown coefficients:

$$U_{PP}(W,Z) = \sum_{i=0}^{2} \sum_{j=0}^{2} C_{i,j} W^{j} (Z - 100)^{i}$$

The initial conditions used to estimate the coefficients were similar to the following. Their reasonableness can be verified by inspection of Figure 5:

(1) $U_{PP}(100,0) = 50$.

(2) $U_{PP}(100,100) = 100$.

(3) $\frac{\partial U_{PP}}{\partial Z} = 0$ when evaluated at $Z = 100$, $W = 100$.

(4) $U_{PP}(0,100) = 90$.

(5) $U_{PP}(0,0) = 0$.

(6) $\frac{\partial U_{PP}}{\partial Z} = 0$ when evaluated at $Z = 0$, $W = 0$.

(7) $U_{PP}(80,100) = 96.4$

(8) $U_{PP}(80,0) = 32$.

(9) $\frac{\partial^2 U}{\partial Z^2} = 0$ when evaluated at $W = 80$.

Condition (9) states that $U(80, Z)$ is a linear function of Z, and so its second partial derivative with respect to Z should equal zero when evaluated at $W = 80$. The standardized Priority/Preference payoff is obtained from the equation:

$$U^*_{PP}(W_j, Z_i) = 50 + 10 \left(\frac{U_{PP}(W_j, Z_i) - \mu_{PP}}{\sigma_{PP}} \right) \text{, where}$$

$U^*_{PP}(W_j, Z_i)$ is the standardized priority/preference payoff associated with individual i and rating j,

$U_{PP}(W_j, Z_i)$ is the unstandardized priority/preference utility for individual i and rating j, and

μ_{PP} and σ_{PP} are the mean and standard deviation, respectively, of $U_{PP}(W_j, Z_i)$ scores in the reference population.

CLASP Parameter Update Considerations

The Navy priority indices, μ_{pp}, and σ_{pp} for each rating are the three Priority/Preference component parameters that require updates. The CLASP parameter update software automatically generates updates for μ_{pp} and σ_{pp} during the annual CLASP parameter update. However, no known documentation describing procedures, methodology, or software for updating the Navy priority indices exists, other than the summary description given in Kroeker and Rafacz and repeated above. The priority indices were never updated during the author's association with CLASP between 1980 and 1999. In the absence of detailed information to supplement the summary description, it is not feasible to perform future Navy priority index updates.

Minority-Fill Component

Prior to CLASP, minority group members were assigned in disproportionately large numbers to a few ratings and in small numbers to many others. The minority-fill component was designed to provide a uniform assignment of minority group members for each rating. A uniform rate of non-minority assignments is also implied. The goal was for the proportion of minority group members in any rating to always equal the previously specified minority proportion goal for the rating.

Kroeker's methodology for determining minority proportion goals during his tenure on the CLASP project is largely undocumented. However, each goal was apparently constructed to compensate for historical minority fill trends. If historical minority fill rates for a given rating were less than historical minority fill rates across all ratings (e.g., Navy-wide minority fill rates), then a minority fill goal larger than the historical average was specified. Conversely, if the historical fill rate for a given rating was greater than the average historical rate across all ratings, then a minority fill goal smaller than the historical average was specified. Beginning with the 2001 CLASP parameter update, NPRST began using a common (Navy-wide) minority goal for all ratings.

Differences between the actual and desired minority group proportions at any given time in the reservation cycle indicate the current status of the uniform fill-rate objective function. The function compensates for current conditions by (1) adding utility points for minority group members and subtracting utility points for non-minority group members when the current proportion of minority group members is less than the minority goal, and (2) subtracting utility points for minority group members and adding utility points for non-minority group members when the current proportion of minority group members is greater than the minority goal. The equation defining the feedback function is given by

$M_{i,j} = (G_j - F_{j,t})I_{M/NM}$, where:

$M_{i,j}$ is the minority fill difference associated with assigning individual i to rating j at time t,

G_j is the desired minority-fill goal for rating j,

$F_{j,t}$ is the actual minority fill proportion for rating j at time t (i.e., the ratio of the number of minority accessions in rating j to the total number of accessions in rating j), and

$I_{M/NM}$ is a variable whose value is 1 if the individual being classified at time t is a minority group member and is -1 if the individual is a non-minority group member.

The standardized minority-fill payoff is computed according to the equation below. The quantity of utility points added or subtracted is proportional to the difference between the actual and desired fill proportions:

$$U_{MF}(i, j, t) = 50 + 10\left(\frac{M_{i,j}}{\sigma_{MF}}\right), \text{ where}$$

$U_{MF}(i, j, t)$ is the standardized minority fill payoff for individual *i* being classified at time *t* with respect to rating *j*,

$M_{i,j}$ is defined in the previous equation, and

σ_{MF} is the standard deviation of $M_{i,j}$ differences in the reference population.

The above equation represents the minority fill payoff function used in the CLASP simulation model and the operational CLASP model. Note that there is a difference in the denominators of this equation and the corresponding equation in Kroeker and Rafacz (1983); the reason for this difference is unknown.

Minority Fill Parameter Update Considerations

The only Minority Fill component parameter subject to updating is σ_{MF}. The CLASP parameter update software automatically generates an update for σ_{MF} during the annual CLASP parameter update.

Fraction Fill Component

Prior to CLASP, the end of each recruiting month was typically marked by a flurry of recruiting activity aimed at filling a substantial number of positions in certain ratings. From a managerial perspective, a procedure resulting in a uniform rate of assignment across all ratings is highly desirable. The fraction fill component was designed to compare the proportion of applicants assigned to a particular rating with the average proportion of applicants assigned to all ratings at the time. If the fill proportion for the rating in question is less than the average fill proportion, additional utility points are awarded to influence the applicant to select the rating. If selected, the rating fill proportion moves closer to average fill rate. Similarly, utility points are subtracted when the proportion of the recruiting goal that has been filled in a given rating exceeds the average fill proportion. If the applicant selects a different rating, the resulting average fill rate increases slightly, thereby moving closer to the rating fill proportion. The operational part of the fraction-fill utility function is given by:

$T_{j,t} = B_t - F_{j,t}$, where

$T_{j,t}$ is the difference in proportions for rating *j* when individual *i* is classified at time *t*,

B_t is the average fill proportion across all ratings at time *t*, and

$F_{j,t}$ is the proportion of applicants that have been assigned to openings with rating *j* up to time *t*.

The standardized fraction fill payoff is calculated as:

$$U_{FF}(i, j, t) = 50 + 10\left(\frac{T_{j,t}}{\sigma_{FF}}\right), \text{ where}$$

σ_{FF} is the standard deviation of $T_{j,t}$ differences in the reference population.

Fraction Fill Parameter Update Considerations

The only Fraction Fill component parameter subject to updating is σ_{FF}. The CLASP parameter update software automatically generates an update for σ_{FF} during the annual CLASP parameter update.

Attrition Component

One apparent motive for adding the Attrition Component to the CLASP model was to incorporate additional non-ASVAB information into the classification process. The "Attrition" concept has a broader definition in context of the Attrition Component than it does in the context of the School Success component. In School Success, attrition is defined solely in terms of "A" School attrition, while in the context of the Attrition Component; it is defined in terms of Navy-wide attrition. The person attribute is the Success Chances of Recruits Entering the Navy (SCREEN) score, which is based upon AFQT, education credential status, and age. Thus, the Attrition component incorporates non-ASVAB information about the applicant's education credential status and his/her age, and information concerning attrition in the rating from sources other than "A" School into the classification process.

Like the Aptitude/Complexity and Navy Priority/Personnel Preference Components, the Attrition component uses an individual characteristic measure and a rating characteristic measure to evaluate utility. The Attrition Component evaluates the utility of assigning a given individual to a given rating, based upon the probability of surviving the first term of enlistment and the attrition severity index (ASI) of the rating. The person characteristic measure is the SCREEN table (Lockman, 1977) and the rating characteristic measure is the ASI. The SCREEN score, which is based upon the individual's education credential status, AFQT score, and age, reflects the probability of successfully completing the first term of his enlistment. The ASI was developed using 5 factors: retention rate, personnel replacement costs, rating size (number of personnel in the rating), rating requirements (need for trained personnel in the rating), and priority (relative importance of the rating) to the Navy. A multiplicative, multi-attribute model was then used to calculate the ASI from the 5 factors (Thomas, Elster, Euske, & Griffin, 1984).

The following discusses the construction of the attrition component policy function describing the utility of assigning an individual to a rating on the basis of the individual's attrition risk and the attrition severity characteristics of the ratings. The SCREEN table is constructed so that an individual is a low (high) risk to attrite during the first term of his enlistment if his SCREEN score is high (low). Accordingly, for purposes of deriving the attrition policy function, the low attrition risk individual is defined as having a SCREEN of 96, while the high attrition risk individual is defined as having a SCREEN of 70. The ASI scale is constructed so that a rating is characterized by high (low) attrition severity if its ASI is large (small). Accordingly, a rating with a low attrition severity problem is defined as having an ASI of 10, while a rating with a high attrition severity problem is defined as having an ASI equal to 80.

SCREEN rank-orders the applicant population and the ASI scale rank-orders the ratings, thus the assignment of a low-risk applicant (high SCREEN) to a rating with a large ASI is a desirable outcome and should receive high utility. In fact, the policy function was constructed so that this assignment received the largest possible value (100). Although the low-risk applicant is also a low risk to attrite from a low ASI rating, it is more sensible from a classification policy standpoint to assign this applicant to the high ASI ratings, and fill the low ASI ratings with individuals characterized by a slightly larger risk to attrite. Accordingly, the assignment of the low-risk applicant to the low-risk rating received an intermediate value of 60. The assignment of a high-risk applicant to a low-risk rating received a value of 55, slightly less than the value of the assignment of the low-risk applicant to the low-risk rating. Finally, the assignment of a high-risk applicant to a high ASI rating results in the largest possible risk that the applicant will attrite. Accordingly, this undesirable outcome received the lowest possible value (0). Substitution of theses four functional specifications yields four linear equations in four unknown coefficients: $C_{0,0}$, $C_{0,1}$, $C_{1,0}$, and $C_{1,1}$.

$U_{Atr}(S, V) = C_{0,0} + C_{1,0}\,(S - 70) + C_{0,1}\,(V - 80) + C_{1,1}\,(S - 70)\,(V - 80)$, where

$U_{Atr}(S, V)$ = non-standardized attrition component utility of assigning person *i* to job option *j*,

S = applicant's SCREEN score,

V = attrition severity index

Solution of the 4 equations yields these estimates: $C_{0,0} = 0.0$, $C_{1,0} = 3.846$, $C_{0,1} = -0.7857$, and $C_{1,1} = 0.0522$.

In Figure 6, the non-standardized attrition utility $U_{Atr}(S, V)$ is plotted on the vertical axis against SCREEN on the horizontal axis, for fixed ASI values of 10, 45, and 80. The standardized Attrition payoff is obtained from the equation:

$$U^*_{Atr}\left(V_j, S_i\right) = 50 + 10\left(\frac{U_{Atr}\left(V_j, S_i\right) - \mu_{Atr}}{\sigma_{Atr}}\right)$$, where

$U^*_{Atr}\left(V_j, S_i\right)$ = standardized attrition component payoff associated with individual *i* and rating *j*,

$U_{Atr}(V_j, S_i)$ = non-standardized attrition component payoff for individual *i* and rating *j*, and

μ_{Atr} and σ_{Atr} are the mean and standard deviation, respectively, of $U_{Atr}(V_j, S_i)$ scores in the reference population.

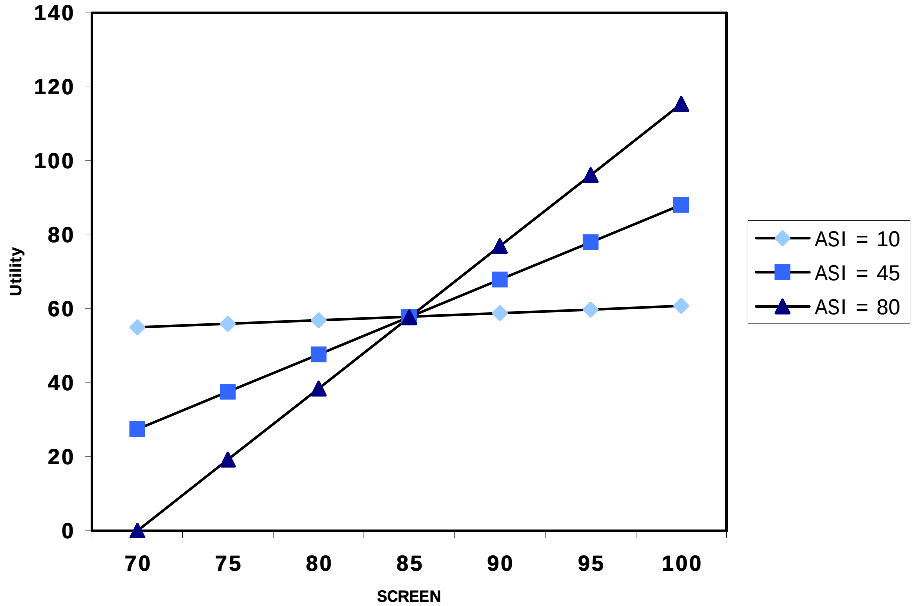

Figure 6. Attrition Utility at constant severity values.

CLASP Parameter Update Considerations

The attrition severity index (ASI) parameters for each rating along with μ_{Atr} and σ_{Atr} constitute the attrition component parameters subject to updating. The CLASP parameter update software automatically generates updates for μ_{Atr} and σ_{Atr} during the annual CLASP parameter update. Thomas, Elster, Euske, and Griffin (1984) document the procedures and methodology they used to develop the original set of ASI parameters in the early 1980s. However, the ASI parameters have not been updated since their 1983 implementation. In the absence of (a) detailed information to supplement the Thomas et al. report, (b) knowledge of and access to all relevant attrition, replacement cost, and demand for personnel information, and (c) software to calculate the updates, it is not feasible to perform future ASI parameter updates.

CLASP Component Weights and Composite Payoff

As previously described, a payoff vector for each CLASP component is calculated, the *j*th entry of which is the standardized payoff of assigning a given individual to the *j*th rating. For each rating, the weighted sum of the six components represents the composite (overall) utility associated with assigning the individual to that rating. The weighted sum for each rating, hereafter called the "composite payoff" is given by:

$$U_{i,j} = \sum_{k=1}^{6} w_k U^{*}_{i,j,k} \quad \text{where} \quad \sum_{k=1}^{6} w_k = 100 \qquad \text{Composite Payoff}$$

where $U_{i,j}$ is the composite payoff for the *i*th individual with respect to job option *j*,

$U^{*}_{i,j,k}$ is the standardized component *k* payoff for the *i*th individual with respect to job option *j*, and w_k is the weight associated with component *k*.

The component weights were determined by Navy classification policy. Each weight expresses, in some sense, the policymaker's desired "contribution" of each component to the composite. In practice, however, "contribution" is difficult to define mathematically. Correlations among the 6 components make it difficult to state an exact relationship between the component weight and the proportion of variance that the component contributes to the composite payoff. However, standardization of the composite payoffs allows CLASP to partially control each component's contribution to the composite variance. As a result, each component weight provides a reasonable approximation to the policymaker's desired contribution of each component.

As described by Kroeker and Rafacz (1983), the component weights were derived according to the following criteria: The raw utility scores for the school success and aptitude/complexity components were examined. It was observed that the variance of the aptitude/complexity scores was affected by a number of extreme values. For the center of the scale to function effectively in discriminating between persons, it was decided that the variance of the weighted aptitude/complexity component should be allowed to assume a larger value than that of the weighted school success component, but by no more than a ratio of 3:2. Respective weights of 26 and 35 for the school success and aptitude/complexity components satisfied this criterion. The second criterion stipulated that the priority/preference component should carry approximately the same weight (14) as the combined minority and fraction file component weights (15). The minority-fill component was given a slightly larger weight than the fraction-fill component, resulting in weights of 8 and 7 respectively. The attrition component weight (10) was assigned according to the requirement that it not exceed the individual weights of the school success, aptitude/difficulty, and priority/preference components.

Table 2
Component weights

Component	Weight
School Success	26
Aptitude/Complexity	35
Priority/Preference	14
Minority-Fill	8
Fraction-Fill	7
Attrition	10

CLASP Decision Indices and Optimality Indicators

This section describes computation of applicants' decision index (DI) and optimality indicator (OI) distributions from their composite payoff vector. CLASP computes the decision index for each rating as the difference between the composite payoff and the corresponding decision index mean:

$$\Delta_{i,j} = U_{i,j} - \overline{U}_j \qquad \text{Decision\_Index}$$

where $\Delta_{i,j}$ is the DI for the ith individual with respect to job option j,

$U_{i,j}$ is the composite payoff for the ith individual with respect to job option j, and

$\overline{U}_j$ is the DIM for job option j.

As previously described, CLASP attempts to force the classifier and applicant to select a job option close to the top of the optimal list and makes it more difficult to select an option near the bottom. However, the joint distribution of the vector of composite utility functions (across all job options) may be such that certain job options make infrequent appearances near the top of the optimal list and, consequently, classifiers cannot access them frequently enough to satisfy recruiting goals. Such a scenario may occur, for instance, when the quota is large and the expected value of the composite utility function for that job option is small, relative to the other job options. To compensate, a decision index mean (DIM) for each job option is subtracted from the applicant's composite utility score for than job. Each DIM is the mean of the composite payoff distribution for that job option with respect to the applicant population (Ward, 1958). For each job, the expected value of the difference between the composite utility and the DIM is zero. This adjustment insures that, over the long run, each job option is as likely to appear near the top of the optimal list as it is to appear near the bottom. Analysis of historical CLASP transaction data has demonstrated that this adjustment is usually adequate to insure that sufficient CLASP presentations are generated to allow classifiers to cover the quota for each job.

The decision indices are transformed onto a scale ranging from 0 to 100 for presentation to the classifier and applicant. This is accomplished in two stages. Stage 1 transforms the individual's decision index distribution onto a first-stage OI scale having a mean of 50 and standard deviation of 20. CLASP performs this transformation using a weighted mean and standard deviation of the individual's DI distribution, where each weight is the current number of available openings (i.e., quota minus reservations-to-date) for the rating and ship month. (Note: Equation [16] in Kroeker and Rafacz [1984] is not consistent with the Stage 1 transformation performed in CLASP. The operational CLASP implementation includes a [weighted] mean, while equation [16) does not.) The Stage 1 OI list is then sorted in descending order. OIs on the sorted list are then translated and truncated to generate the Stage 2 list. The combined translation and truncation operations give the highest-rated rating on the Stage 2 scale an OI of 100 and insure that none of the OIs at the bottom of the list are less than zero. Equation (17) of Kroeker and Rafacz (1984) describes the translation. After translation, each negative OI on the list is set equal to zero.

RIDE Composite Payoff

The RIDE composite utility for individual *i* and job option j is

$$C_{i,j} = W_{SPSU} S_{i,j}^{III} + W_{AFQT} Q_{i,j}$$

where $W_{SPSU} = W_{AFQT} = 1/2$ are the respective SPSU and AFQT component weights,

$C_{i,j}$ is the composite RIDE utility for individual *i* and job option *j*, and

$S_{i,j}^{III}$ and $Q_{i,j}$ are the SPSU (Stage III) and AFQT utility scores for individual *i* and job option *j*.

Discussion

This section discusses certain issues raised during the course of the CLASP-RIDE comparison that may be relevant to classification policymakers. These issues include (1) standardization of RIDE components, (2) incorporation of factors into the classification decision that are excluded from the CLASP and RIDE algorithms, including non-psychological/psychometric variables, and certain dynamic and time-critical factors, (3) the PDR concept and parameterization of RIDE, and (4) discussion of Bin model vs. LRM results.

CLASP standardizes each of its 6 components so that each has a mean of 50 and standard deviation of 10. As previously described, CLASP policymakers apparently felt it was important to apply appropriate nominal weights to each component and standardize the component score distributions in such a manner that the effective component weights closely approximate the nominal weights. If classification policymakers are also concerned about consistency between the nominal and effective weights of the SPSU and AFQT components, they should recognize that the nominal weights for the SPSU and AFQT components (currently 50% for each) are probably not the same as the effective weights. The actual weight of each RIDE component is determined by the product of the nominal weight and the standard deviation of the component. Hence, the component with the largest standard deviation has the largest effective weight. If classification policymakers wish to specify the actual weight of each component, each component must be standardized (by dividing by its standard deviation) before calculation of the RIDE composite payoff.

In the current CLASP implementation, classifiers must attempt to sell each applicant an option from the Top 15 prior to viewing the CLASP optimal list. However, before introduction of the Top 15 feature, the CLASP optimal list presentation strategy indicates there was considerable emphasis on placing each applicant into an "optimal" job assignment. This emphasis was manifested in the manner in which the optimality output was used in the classification interview, particularly in the optimal list presentation strategy and the classifier's role in selling an option on the list. This strategy forced the classifier and applicant to view the CLASP optimal list in groups of 5, 10, or 15 job options at a time, depending upon the applicant's projected enlistment date. The first group of options consisted of those jobs with the highest optimality scores. In theory, the classifier's role was to convince the applicant to buy an option on this list because they were considered the best possible matches. If the classifier could not sell one of these options, he would try to sell an option from the job group with the next largest set of optimality scores. In theory, the classifier would continue working down the optimal list until a group containing a mutually satisfactory option was found. Although it was possible for classifiers to access and sell options near the bottom of the CLASP optimal list, it was more difficult and time consuming for them to do so.

CLASP was developed during a period when several papers in the Industrial/Organizational psychology literature touted the potential benefits of automating empirical models describing the utility of matching applicants to jobs (Dunnette & Borman, 1979). An attitude prevailed that most or all factors considered during personnel classification decisions could and should be implemented on the computer. In apparent accordance with this point of view, CLASP was sold to Commander, Navy Recruiting Command (CNRC) under the philosophy that a computerized classification algorithm could rank order job options by their mutual benefit to both the Navy and applicant. The presentation strategy described above clearly promotes CLASP's definition of optimality by reinforcing the classifier to select from the top of the list.

Enlisted recruit classification occurs in an environment that places a strong emphasis on filling quotas and meeting recruiting requirements and objectives. The classification algorithm must operate in a manner consistent with the attainment of these goals. These goals originate outside of CNRC. Some Navy ratings have large recruiting goals, due to large manpower requirements, while other ratings have comparatively small goals. Some ratings are popular and comparatively easy to sell, while others are less popular and more difficult to sell. Changes in recruiting goals and shifting of quotas among different recruiting cycles occur frequently. Changes in the Navy's perception of which recruiting goals are critical in nature often occur. The events that precipitate changing recruiting goals and changing criticality designations may be difficult to forecast in advance. Hence, the dynamic nature of the operational Navy environment means that designations of which jobs are considered critical, their relative degrees of criticality, and recruiting goals can change suddenly and without warning.

CLASP is unable to fill job quotas evenly and re-channel applicants into critical jobs without substantial classifier intervention. Although the purpose of the Fraction Fill component is to fill quotas evenly across job options, it has minimal impact on achieving this because it carries only 7 percent of the weight in the CLASP optimality composite. CLASP cannot re-channel applicants into critical jobs because it cannot differentiate between jobs on the basis of their criticality. Instead, its definition of optimality focuses on the psychological measures of goodness of fit between person and job. The inputs to these functions are those person and job characteristics that are relatively stable, permanent, and enduring in nature. These include factors such as intelligence, job-specific aptitude, and job technical complexity. In short, it is not possible for CLASP or RIDE to adjust for all factors that should be included in the classification process. Classification algorithms such as RIDE and CLASP can optimize their assignment recommendations based only on the more permanent and enduring characteristics of person and job, in particular, the psychological/psychometric variables they currently use. Classifiers must override classification algorithm recommendations if they desire to incorporate the more dynamic and time-critical factors into the classification decision. Therefore, classification algorithms and their optimal list presentation strategies must give the classifier a convenient way to sell any job currently experiencing a critical need, regardless of that job's ranking on the classification algorithm's optimal list. In CLASP, the use of the decision index to rank-order the job options (instead of the composite payoff) has helped insure that all ratings are reasonably accessible to classifier and applicant, even when the classifier was expected to sell from the top of the optimal list.

In contrast, RIDE does not employ the DIM concept. It rank-orders job options on the basis of RIDE composite utility. This may be entirely valid. RIDE is being developed under different user expectations than CLASP was. Unlike CLASP, RIDE is being implemented on modern hardware. The DIM concept may be completely unnecessary in RIDE if user expectations and hardware capabilities are such that RIDE can provide adequate accessibility to all ratings, regardless of optimality value.

Parameterization of RIDE Model

One attractive feature of CLASP is that “A” School student performance data is not required to parameterize the model. The same is not true for RIDE. Student performance data for each RIDE job option is required to both find the PDR and estimate the FPPS rates in the cut score and PDR bins. However, in the Bin Model Evaluation section, it was demonstrated that the PDR concept does not stand up to rigorous statistical testing, except in a small number of ratings.

Given the following problems associated with the use of school performance data to parameterize RIDE, it is reasonable to ask whether the RIDE model concept should be modified to eliminate the need for school performance data from the parameter update process. These include:

- Weak empirical support for the current PDR concept
- Inexperience and uncertainty with respect to the FPPS criterion and data sources
- School performance data is not necessary to parameterize RIDE, except to update the PDRs
- Collection of school performance data for PDR update purposes would require time, money, and effort far beyond that required to collect only applicant data for the same purpose
- In some cases, ASVAB selector composite and/or cut score changes cannot be implemented immediately in RIDE, due to the inappropriateness of using an outdated validation sample to update a PDR parameter

The design of the SPSU and AFQT components depends heavily on the validity of the PDR concept. If one considers the current PDR concept to be invalid, but still believes the SPSU and AFQT components to be valid mathematical models of the goodness-of-fit between person and job, then an alternative PDR concept is needed and an alternative procedure is needed to estimate the PDRs. The alternative procedure must not depend on a hypothesized empirical relationship between FPPS and student aptitude. In addition, the procedure should be constructed so that parameters can be estimated from Navy applicant data only. School performance data should not be required to estimate the parameters.

In the author's opinion, the RIDE algorithm can be justified as a classifier decision process model and the PDR can be justified as an important parameter in that model. An estimation procedure satisfying these requirements can then be derived from the concept of RIDE as a classifier decision model. Suppose a classifier, without assistance from an automated classification algorithm such as CLASP or RIDE, must classify an applicant. Suppose the applicant satisfies the cut score in each option being considered. Suppose the classifier knows (a) the applicant's composite and AFQT scores, (b) cut scores for all composites, and (c) all composite score distributions relative to the applicant population. A simple mathematical model can be developed to classify the applicant based on the given information. The model is based on the assumption that the classifier uses reference points on the composite and AFQT score distributions as rules-of-thumb for determining which job option the applicant is best suited for.

In this model, the classifier uses one such reference point in the same manner as a PDR, that is, to designate a decision cut-off point which he may use to determine whether his applicant is marginally qualified, maximally qualified, or over-qualified for a given job. If the composite score exceeds the PDR, then the classifier may consider the applicant as either over-qualified for the job under consideration (and thus a potential candidate for a more difficult job) or maximally qualified (and thus a solid candidate for the job under consideration). If the composite score is less than the PDR, then the classifier may consider the applicant as marginally qualified for the job and, therefore, a potential candidate for a less difficult job. As previously described, the AFQT component decides whether the applicant is over-qualified or maximally qualified. The decision currently depends upon where the applicant's AFQT score stands in relation to the over-qualification point (M + Delta) on the AFQT distribution for that job option.

Future research is required to further analyze and fill in the details of this proposed modification. Unanswered issues remain concerning the size of the marginally-qualified, maximally-qualified, and over-qualified regions. Should they all be about the same size, in terms of the proportions of the applicant population residing in each region? Depending upon the answer, consideration should be given to modifying the AFQT component's current decision rule.

As this report is being finalized, it is uncertain whether "A" School performance data will be available and whether its use will be feasible for input to the RIDE parameter update process. This report has also raised questions concerning the lack of empirical support for the PDR concept and the Navy's lack of experience with and understanding of FPPS. If school performance data is either unavailable or infeasible for use, then questions regarding the appropriateness of the Bin model to estimate FPPS are irrelevant. As previously discussed, it will be necessary to reformulate the RIDE model in terms of some underlying concept other than PDR as an indicator of student over-qualification.

However, if it is determined that school performance data is available and feasible for use (and the FPPS criterion and original PDR concept is still considered valid), then the appropriateness of the Bin model for FPPS estimation purposes becomes an important issue. In particular, NPRST must then determine what FPPS estimation methodologies may be more appropriate and more accurate than the Bin procedure. The results in Table 1 strongly suggest that the LRM is superior to the Bin procedure, particularly when both bias and estimation error are taken into consideration. In addition, from a mathematical and software implementation standpoint, the LRM is no more complex than the Bin model.

Conclusions and Recommendations

If classification policymakers desire that the nominal and effective weights of the RIDE components remain approximately equal, RIDE must be modified so that each component is standardized (by dividing by its standard deviation) before calculating the RIDE composite payoff.

Assume that (1) classifiers using RIDE are not under any obligation to sell from the top of the optimal list and (2) unlike CLASP, there are no constraints on classifier access to the lower portions of the RIDE optimal list and it is equally convenient for him to sell a job from the bottom of the list as it is for him to sell one from the top. Then, the DIM concept is not required in the RIDE model because a classifier using RIDE has sufficient freedom to put the applicant into a job option with a low optimality value if quota fill and/or criticality requirements dictate that he do so.

If “A” School performance data is unavailable or is determined to be infeasible for use in the RIDE parameter update (or the current PDR concept is considered invalid), then classification policymakers should consider the classifier decision model as a potential alternative for redefining the PDR concept and becoming the conceptual framework for the RIDE parameter update process.

If school performance data is available and feasible for use in the RIDE parameter update, then classification policymakers should consider the LRM as a replacement for the Bin methodology.

References

Cronbach, L. T. & Gleser, G. C. (1965). *Psychological testing and personnel decisions*. Urbana: University of Illinois Press.

Dunnette, M. D. & Borman, W. C. (1979). Personnel selection and classification systems, *Annual review of psychology*. 30, 477-525.

Folchi, J. S. (1999). Specifications for RIDE utility functions (unpublished).

Hannan, J. F. & Tate, R. F. (1965). Estimation of the parameters for a multivariate normal distribution when one variable is dichotomized. *Biometrika*, 52, 664-668.

Hosmer, D. W. & Lemeshow, S. (1989). *Applied logistic regression*. New York: John Wiley and Sons

Kroeker, L. K. & Folchi, J. S. (1984). *Classification and assignment within PRIDE (CLASP) system: Development and evaluation of an attrition component*. (NPRDC-TR-84-40). San Diego: Navy Personnel Research and Development Center.

Kroeker, L. K. & Rafacz, B. A. *Classification and assignment with PRIDE (CLASP): A recruit assignment model*. (NPRDC-TR-84-9). San Diego: Navy Personnel Research and Development Center.

Kroeker, L. K. (1982). A procedure to revise estimates of psychological scale values. In B. Rimland (ed.), *Independent research and independent exploratory development at the Navy Personnel Research and Development Center--FY81*. (NPRDC Spec. Rep. 82-77). San Diego: Navy Personnel Research and Development Center.

Lockman, R. F. (1977). *Success chances of recruits entering the Navy (SCREEN)* (No. 1086) Arlington , VA: Center for Naval Analyses.

Prince, B. M. & Tate, R. F. (1966). Accuracy of maximum-likelihood estimates of correlation for a biserial model. *Pschometrika*, 31, 85-92.

Rodin, B. (1970). *Calculus with analytic geometry*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Stuart, A., & Ord, J. K. (1991). *Kendal's Advanced Theory of Statistics*. New York: Oxford University Press.

Thomas, G., Elster, R., Euske, K., & Griffin, P. (1984). *Attrition severity index (ASI) for selected Navy ratings: Development*. (NPRDC-TR-85-1). San Diego: Navy Personnel Research and Development Center.

Ward, Joe H. Jr. (1977). *Creating mathematical models of judgment processes: From policy-capturing to policy-specifying*. (AFHRL-TR-77-47). Brooks Air Force Base, TX: Air Force Human Resources Laboratory1977.

Wonnacott, T. H. & Wonnacott, R. J. (1972). *Introductory Statistics for Business and Economics*. NewYork: John Wiley & Sons, Inc.

Appendix A:
Derivation of EATE Formula

Derivation of EATE Formula and Pseudo Code to Compare the Bin and LRM Models

Total Error $\varepsilon \sim N(\mu, \sigma^2)$, $\mu = \text{bias}$, $\sigma^2 = \text{estimation error variance.}$

Define $\varphi(t) =$ standard normal prob. density function,

Define $\Phi(t) =$ Standard normal cumulative distribution function

$$\text{EATE} \;=\; E|\varepsilon| = \int_{-\infty}^{\infty} |\,t\,| \frac{1}{\sigma} \varphi\left(\frac{t-\mu}{\sigma}\right) dt$$

$$\text{EATE} = \int_{-\infty}^{0} -t \frac{1}{\sigma} \varphi\left(\frac{t-\mu}{\sigma}\right) dt + \int_{0}^{\infty} t \,\frac{1}{\sigma} \varphi\left(\frac{t-\mu}{\sigma}\right) dt$$

Substitute $y = \dfrac{t-\mu}{\sigma}$ and $dy = \dfrac{1}{\sigma} dt$ with $t = \sigma\, y + \mu$

$$\text{EATE} = \int_{-\infty}^{-\frac{\mu}{\sigma}} \text{-}(\sigma\,\mathrm{y} + \mu)\varphi(y)dy + \int_{-\frac{\mu}{\sigma}}^{\infty} (\sigma\, y + \mu)\varphi(y)dy$$

$$\text{EATE} \;=\; \left[\sigma\,\varphi(\mathrm{y}) - \mu\,\Phi(y)\right]_{-\infty}^{-\frac{\mu}{\sigma}} + \left[-\,\sigma\varphi(y) + \mu\,\Phi(y)\right]_{-\mu/\sigma}^{\infty}$$

$$\text{EATE} = 2\sigma\,\varphi\left(-\frac{\mu}{\sigma}\right) - 2\mu\,\Phi\left(-\frac{\mu}{\sigma}\right) + \mu$$

Pseudo Code to compare the Bin and LRM models:

Outer Loop over j = 1, 70 RIDE job options:

Inner Loop over composite score $X_{i,j}$, where $CS \le X_{i,j} \le C_{Max}$.

`**LRM Calculation**: Calculate LRM EATE for current value of $X_{i,j}$.
Assume Bias in Logit = 0, and Variance of logit is computed as follows:
The estimated logit $\hat{L}(X_{i,j})$ is

For LLRM: $\hat{L}(X_{i,j}) = \sum_{k=0}^{1} \hat{\beta}_{k,j} X_{i,j}^{k}$ For QLRM: $\hat{L}(X_{i,j}) = \sum_{k=0}^{2} \hat{\alpha}_{k,j} X_{i,j}^{k}$

The variance of the estimated logit is given by the quadratic form:
$(1 \quad \vec{x}^{\,t})\hat{\Sigma}(1 \quad \vec{x}^{\,t})^t$ where $\hat{\Sigma}$ is the estimated covariance matrix of the parameter estimates, $\vec{x} = \begin{pmatrix} 1 \\ X_{i,j} \end{pmatrix}$ for the LLRM, $\vec{x} = \begin{pmatrix} 1 \\ X_{i,j} \\ X_{i,j}^2 \end{pmatrix}$ for the QLRM, and the superscript t indicates matrix transposition. Then, calculate EATE in Logit by substituting the logit

bias = 0 and the logit variance into equation EATE to obtain the EATE of the logit ($EATE\left(\hat{L}\left(X_{i,j}\right)\right)$ in LRM EATE formula):

Calculate EATE of LRM FPPS rate estimate (LRM EATE) for $X_{i,j}$ by

$$\frac{1}{1+\exp\left[-\left(\hat{L}\left(X_{i,j}\right)+EATE\left(\hat{L}\left(X_{i,j}\right)\right)\right)\right]}-\frac{1}{1+\exp\left[-\left(\hat{L}\left(X_{i,j}\right)-EATE\left(\hat{L}\left(X_{i,j}\right)\right)\right)\right]}$$

For current job option j, accumulate sum of LRM EATE over all $X_{i,j}$.

Bin Calculation: Calculate EATE(Bin) = EATE of Bin model FPPS rate estimator:

For current X, compute Bin bias = difference between Stage I and Stage II Models at the current value of $X_{i,j}$. The stage I estimator is:

$$S_{i,j}^{I}\left(X_{i,j}\right)=\frac{\hat{F}_j(m)-\hat{F}_j(m-1)}{b_j(m)-b_j(m-1)}\left(X_{i,j}-b_j(m-1)\right)+\hat{F}_j(m-1)$$

where $\hat{F}_j(m)$ is the FPPS rate in the *m*th bin of school sample *j*, and $X_{i,j}$ is located between the midpoint of the (*m-1*)th and *m*th bins, i.e. $b_j(m-1)\le X_{i,j}\le b_j(m)$.

For current X, calculate stage I model estimation error variance for the current $X_{i,j}$ satisfying $b_j(m-1)\le X_{i,j}\le b_j(m)$. The variance is

$$\left\{\frac{X_{i,j}-b_j(m-1)}{b_j(m)-b_j(m-1)}\right\}^2\mathrm{Var}\left(\hat{F}_j(m)\right)+\left\{\frac{b_j(m)-X_{i,j}}{b_j(m)-b_j(m-1)}\right\}^2\mathrm{Var}\left(\hat{F}_j(m-1)\right)$$

where $\mathrm{Var}\left(\hat{F}_j(m)\right)=\dfrac{\hat{F}_j(m)\left(1-\hat{F}_j(m)\right)}{N_j(m)}$

The variance is derived by computing the variance of the stage I estimate, and using the properties of the binomial distribution and the independence of the FPPS rates in the *m*th and (*m-1*)th bins.

For current X, calculate EATE(Bin) using the Bin bias and the stage I estimation error variance and substituting them into equation EATE.

Accumulate sum of EATE(Bin) over all $X_{i,j}$

End Inner Loop ($CS\le X_{i,j}\le C_{Max}$ loop).

For job option *j*, compute LRM_EATE = average of EATE(LRM) = mean EATE of LRM FPPS rate estimates over all $X_{i,j}$ { $X_{i,j}$ | $CS\le X_{i,j}\le C_{Max}$ }.

For job option j, compute Bin_EATE = average of EATE(Bin) = mean EATE of Bin FPPS rate estimates over all $X_{i,j}$: { $X_{i,j}$ | $CS\le X_{i,j}\le C_{Max}$ }.

End Outer Loop (RIDE job option j loop).

Compute Overall_LRM_EATA = average of Expected AbsoluteTotal Error of LRM FPPS rate estimates over all job options *j* and all $X_{i,j}$.

Compute Overall_Bin_EATA = average of Expected AbsoluteTotal Error of Bin FPPS rate estimates over all job options *j* and all $X_{i,j}$.

Appendix B:
Finding the QLRM Extreme Value Point

Finding the QLRM Extreme Value Point

Show that QLRM has exactly one extreme value point and find it. Show that the extreme value point is a minimum if $\alpha_{2,j} > 0$ and is a maximum if $\alpha_{2,j} < 0$: The QLRM is given by:

$$S_{i,j}^{Q} = \{1 + \exp[-(\alpha_{2,j} X_{i,j}^{2} + \alpha_{1,j} X_{i,j} + \alpha_{0,j})]\}^{-1}$$

Extreme value point is found by solving $\frac{dS_{i,j}^{Q}}{dX_{i,j}} = 0$ for $X_{i,j}$ (Rodin, 1970). By the chain rule,

$\frac{dS_{i,j}^{Q}}{dX_{i,j}} = \frac{dL}{dy}\frac{dy}{dX_{i,j}}$ where $L(y) = \{1 + \exp[-y]\}^{-1}$ and

$y(X_{i,j}) = \alpha_{2,j} X_{i,j}^{2} + \alpha_{1,j} X_{i,j} + \alpha_{0,j}$.

$\frac{dL}{dy} = \frac{\exp(-y)}{(1+\exp(-y))^{2}} > 0$ for all y and $\frac{dy}{dX_{i,j}} = 2\alpha_{2,j} X_{i,j} + \alpha_{1,j}$.

Hence, $\frac{dS_{i,j}^{Q}}{dX_{i,j}} = 0$ if and only if $2\alpha_{2,j} X_{i,j} + \alpha_{1,j} = 0$. Therefore, solving this equation for $X_{i,j}$,

we see that the only extreme value point occurs at $X_{Ext} = \frac{-\alpha_{1,j}}{2\alpha_{2,j}}$.

Extreme value point X_{Ext} is a minimum iff $\frac{d^{2}S_{i,j}^{Q}}{dX_{i,j}^{2}} < 0$ and is a maximum if and only if $\frac{d^{2}S_{i,j}^{Q}}{dX_{i,j}^{2}} >$

0,

where $\frac{d^{2}S_{i,j}^{Q}}{dX_{i,j}^{2}}$ is evaluated at X_{Ext} (Rodin, 1970).

$\frac{d^{2}S_{i,j}^{Q}}{dX_{i,j}^{2}} = \frac{d^{2}L}{dy^{2}}\left(\frac{dy}{dX_{i,j}}\right)^{2} + \frac{dL}{dy}\frac{d^{2}y}{dX_{i,j}^{2}}$. However, $\frac{dy}{dX_{i,j}} = 0$ at $X_{i,j} = X_{Ext}$, so

$\frac{d^{2}S_{i,j}^{Q}}{dX_{i,j}^{2}} = \frac{dL}{dy}\frac{d^{2}y}{dX_{i,j}^{2}} = 2\alpha_{2,j}\frac{dL}{dy}$. Since $\frac{dL}{dy} > 0$, $\frac{d^{2}S_{i,j}^{Q}}{dX_{i,j}^{2}} < 0$ if and only if

$\alpha_{2,j} < 0$ and $\frac{d^{2}S_{i,j}^{Q}}{dX_{i,j}^{2}} > 0$ if and only if $\alpha_{2,j} > 0$. QED.

Appendix C:
Use of P-value

Use of P-value

In our testing situation, we set a null hypothesis for the QLRM and a null hypothesis for the LLRM. The QLRM null hypothesis is that the X^2 coefficient in the QLRM is zero, while the LLRM null hypothesis states that the X coefficient in the LLRM is zero. Opposing each null hypothesis is the corresponding alternative hypothesis stating the coefficient is non-zero. For two-sided tests such as these, the p-value may be defined as the probability that the test statistic is at least as large in absolute value as the parameter estimate actually observed if the null hypothesis were true. Stated differently, the p-value represents the probability that the experimenter incorrectly rejects the null hypothesis on the basis of his observed parameter estimate. Thus, a small p-value implies small credibility for the null hypothesis and a large p-value implies large credibility for the null hypothesis. Hence, the p-value associated with the X^2 coefficient estimate in the QLRM is a convenient way to measure the credibility of the QLRM null hypothesis, while, the p-value associated with the X coefficient estimate in the LLRM is a convenient way to measure the credibility of the LLRM null hypothesis (Wonnacott & Wonnacott, 1972).

In logistic regression analysis, the test statistic is the square of the ratio of the parameter estimate and its standard estimation error. In our case, the LLRM null hypothesis states that the slope parameter in the LLRM is zero and, therefore, composite score is not useful in predicting FPPS. If the LLRM null hypothesis is false and the LLRM alternative hypothesis $H_{A,LLRM} : \beta_{L,1} \neq 0$ is true, then composite score results in a statistically significant improvement in predicting FPPS. Under the null hypothesis, the square of the ratio of the estimated slope parameter divided by its standard estimation error has a chi-square distribution with 1 degree of freedom. An analogous argument shows that if the QLRM null hypothesis is rejected and the QLRM alternative hypothesis $H_{A,QLRM} : \alpha_{Q,2} \neq 0$ is true, then we may conclude there is statistical evidence that an extreme value point (i.e., maximum or minimum) exists.

Appendix D:
CLASP Aptitude / Difficulty Utility Function Features

Features of Aptitude / Difficulty Utility Function

In Figure 3, $U_{A/D}(A,D)$ is plotted on the vertical axis against Job Difficulty D on the horizontal axis for fixed Aptitude values A = 40, 50, 60, 80, 90, and 99. Equation (Apt_Dif) indicates that this function is both a quadratic in A (for fixed D) and a quadratic in D (for fixed A). Each curve in Figure 3 represents $U_{A/D}(A,D)$ for a fixed A. Since each curve is a quadratic in D, it has exactly one maximum on the job difficulty interval between D=40 and D=99. The maximum, hereafter called $D_{Max}(A)$, occurs at the difficulty level D that awards the largest utility score for the applicant whose aptitude is A. $D_{Max}(A)$ may be obtained by setting the partial derivative of $U_{A/D}(A,D)$ with respect to D equal to zero, then solving for D:

$$\frac{\partial U_{A/D}}{\partial D} = B_{0,1} + 2B_{2,2}(A-100)^2(D-35) + B_{2,1}(A-100)^2 + 2B_{0,2}(D-35) = 0.$$

$$D_{Max} = 35 - \frac{B_{0,1} + B_{2,1}(A-100)^2}{2B_{2,2}(A-100)^2 + 2B_{0,2}}$$

Table D-1 shows the $D_{Max}(A)$ value associated with each Aptitude score between 40 and 99, inclusive. One can verify that $D_{Max}(A)$ maximizes $U_{A/D}(A,D)$ for all D by observing that the 2nd partial derivative of U with respect to D is less than zero for $40 \le A \le 99$.

$$\frac{\partial^2 U_{A/D}}{\partial D^2} = 2B_{2,2}(A-100)^2 + 2B_{0,2} < 0$$

In addition, $D_{Max}(A)$ is monotonically increasing in A. As demonstrated in Table D-1 and mathematically below, both $D_{Max}(A)$ and $U_{A/D}(A, D_{Max}(A))$ are monotonically increasing in A.

Differentiating $\frac{\partial U_{A/D}}{\partial A}$ with respect to A, we obtain

$$\frac{\partial U_{A/D}}{\partial A} = 2(A-100)\left[B_{2,0} + B_{2,1}(D-35) + B_{2,2}(D-35)^2\right]$$

Since $A-100 < 0$, $B_{2,0} < 0$, $B_{2,1} < 0$, and $B_{2,2} < 0$, $\frac{\partial U_{A/D}}{\partial A} > 0$ for all A between 40 and 99, inclusive, and for all D between 40 and 99, inclusive. Thus, for any given D, $U_{A/D}(A_1,D) > U_{A/D}(A_2,D)$ if $A_1 > A_2$. In particular, this is true for D = $D_{Max}(A)$.

Table D-1
$D_{Max}(A)$ value associated with each Aptitude score between 40 and 99

Apt	$D_{Max}(A)$	$U(A,D_{Max}(A))$	Apt	$D_{Max}(A)$	$U(A,D_{Max}(A))$
40	43.1	33.1	70	65.7	55.6
41	43.5	33.4	71	67.0	57.0
42	43.9	33.8	72	68.4	58.4
43	44.3	34.3	73	69.9	59.8
44	44.8	34.7	74	71.4	61.4
45	45.2	35.1	75	73.0	62.9
46	45.7	35.6	76	74.6	64.6
47	46.2	36.1	77	76.3	66.2
48	46.7	36.6	78	78.0	68.0
49	47.2	37.1	79	79.8	69.8
50	47.8	37.7	80	81.6	71.6
51	48.3	38.3	81	83.5	73.4
52	48.9	38.9	82	85.4	75.3
53	49.6	39.5	83	87.3	77.2
54	50.2	40.2	84	89.2	79.2
55	50.9	40.8	85	91.1	81.1
56	51.6	41.5	86	93.1	83.0
57	52.4	42.3	87	95.0	84.9
58	53.2	43.1	88	96.8	86.8
59	54.0	43.9	89	98.6	88.6
60	54.8	44.7	90	100.4	90.4
61	55.7	45.6	91	102.0	92.0
62	56.6	46.6	92	103.6	93.6
63	57.6	47.5	93	105.0	95.0
64	58.6	48.5	94	106.3	96.3
65	59.7	49.6	95	107.4	97.4
66	60.8	50.7	96	108.3	98.3
67	61.9	51.8	97	109.1	99.1
68	63.1	53.0	98	109.6	99.6
69	64.4	54.3	99	109.9	99.9

Appendix E:
Disentangling the SPSU and AFQT Components

Disentangling the SPSU and AFQT Components

The Disentangle program looks at four alternative θ_j (theta) definitions and five alternative applicants. θ_j is where AFQT utility reaches minimum (zero) after declining from its maximum. The alternative θ_j s are given by $M_j + N\sigma_j$, where N = 1.0, 2.0, 3.0, and 3.5. $M_j + \delta_j$ is where AFQT utility begins to decline from its maximum value (100) toward its minimum value of zero. The Disentangle program always defines $M_j + \delta_j$ = $M_j + \frac{1}{2}\sigma_j$.

In this example, we look at one applicant whose AFQT score is 90. N has been fixed at 3.5. Thus, $\theta_j = M_j + 3.5\sigma_j$. The applicant qualified for all jobs (X ≥ CutS). On all jobs for which composite score X exceeded the PDR, $S_{i,j}$ was equal to 100 * Hardness factor (HF). Hence, applicant achieves highest possible $S_{i,j}$ score on all these jobs.

The applicant's AFQT (90) was greater than $M_j + \delta_j$ for all jobs, except NF and MM-NF. Hence, applicant achieves highest possible $Q_{i,j}$ = 100 only on NF and MM-NF. The applicant's AFQT (90) was greater than θ_j only for SKS-SG and SM-SG. Hence, the applicant received the lowest possible $Q_{i,j}$ (zero) on these 2 jobs. Other than MM-NF, NF, SKS-SG, and SM-SG, his AFQT was greater than $M_j + \delta_j$ and was less than θ_j. Therefore, the applicant's $Q_{i,j}$ on these jobs was between zero (the minimum) and 100 (the maximum).

				Sch P/L Suc Util					AFQT Utility				AFQT Utility		RIDE Composite Utility			
				Unsorted			Sorted		Unsorted				Sorted		Unsorted		Sorted	
r	Rate	CutS	X	PDR	HF	Sir	Rate	Sir	Rate	M+Del	Theta	Qir	Rate	Qir	Rate	Cir	Rate	Cir
1	AB GE	130	175	157.5	.12	12	STG GE	82	AB GE	52.3	90.2	1	MM NF	100	AB GE	6	STG GE	83
2	AC 5Y	210	252	247.5	.59	59	FT GE	82	AC 5Y	70.0	103.8	41	NF	100	AC 5Y	50	FT GE	82
3	AD SG	190	245	217.5	.33	33	CTT SG	82	AD SG	56.8	95.6	14	CTI SG	98	AD SG	24	NF	82
4	AE SG	218	245	245.5	.69	69	HT GE	79	AE SG	76.2	109.0	58	ETS GE	85	AE SG	63	EW SG	80
5	AECFAE	218	245	265.5	.69	55	MT AE	76	AECFAE	85.4	111.8	83	EW SG	85	AECFAE	69	MM NF	79
6	AG SG	214	252	236.5	.64	64	EW SG	76	AG SG	71.6	107.7	49	STG GE	84	AG SG	57	CTT SG	77
7	AK SG	103	122	105.5	.54	54	JO 5Y	72	AK SG	68.1	102.0	35	AECFAE	83	AK SG	45	JO 5Y	76
8	AM GE	164	175	181.5	.70	70	CTR SG	72	AM GE	70.4	113.7	55	EA 5Y	83	AM GE	62	CTM AE	73
9	AO SG	190	245	217.5	.33	33	AM GE	70	AO SG	57.7	93.5	10	CTM AE	82	AO SG	22	MT AE	73
10	AS SG	200	252	202.5	.46	46	AE SG	69	AS SG	57.5	93.2	9	FT GE	82	AS SG	28	CTR SG	70
11	AT GE	156	182	193.5	.56	51	RM SG	68	AT GE	78.3	111.1	64	JO 5Y	80	AT GE	57	CTI SG	70
12	AZ SG	103	122	125.5	.54	53	PN SG	67	AZ SG	68.2	103.7	39	EW AE	79	AZ SG	46	ETS GE	70
13	BU 5Y	150	171	157.5	.46	46	CTM AE	64	BU 5Y	62.0	107.3	38	CTT SG	73	BU 5Y	42	EA 5Y	69
14	CE 5Y	196	252	218.5	.41	41	NF	64	CE 5Y	59.5	101.7	28	MT AE	71	CE 5Y	34	AECFAE	69
....																		
44	MM NF	242	245	332.5	1.00	57	UT 5Y	46	MM NF	91.8	114.0	100	UT 5Y	36	MM NF	79	OS SG	42
45	MMS SG	147	182	174.5	.41	41	EO 5Y	46	MMS SG	61.9	98.9	24	AK SG	35	MMS SG	33	BU 5Y	42
46	MN SG	158	171	160.5	.60	60	SW 5Y	46	MN SG	66.7	113.4	50	RP SG	34	MN SG	55	UT 5Y	41
47	MR SG	158	175	185.5	.60	55	AS SG	46	MR SG	66.4	109.3	45	SK SG	33	MR SG	50	YN SG	41
48	MS SG	89	122	116.5	.18	18	HM GE	44	MS SG	53.6	93.8	9	EO 5Y	33	MS SG	14	EN AT	41
49	MSS SG	147	182	164.5	.41	41	DT GE	44	MSS SG	61.0	97.4	20	CE 5Y	28	MSS SG	31	CM 5Y	40
50	MT AE	57	64	69.5	.82	76	CTI SG	43	MT AE	83.0	107.1	71	SW 5Y	28	MT AE	73	EO 5Y	39
51	NF	242	245	337.5	1.00	64	CM 5Y	43	NF	92.2	113.1	100	QM SG	26	NF	82	SW 5Y	37
52	OS SG	157	183	174.5	.58	58	EM SG	41	OS SG	62.9	99.3	26	OS SG	26	OS SG	42	CE 5Y	34
53	PH 5Y	103	122	115.5	.54	54	MMS SG	41	PH 5Y	70.8	108.7	49	EM SG	25	PH 5Y	52	DT GE	34
54	PN SG	108	122	120.5	.67	67	SM SG	41	PN SG	77.3	107.5	58	TM SG	25	PN SG	62	MMS SG	33
55	PR SG	158	171	170.5	.60	60	SS SF	41	PR SG	68.0	112.7	51	MMS SG	24	PR SG	55	TM SG	33
56	QM SG	97	122	119.5	.38	38	TM SG	41	QM SG	60.4	100.2	26	YNS SG	24	QM SG	32	EM SG	33
57	RM SG	163	183	170.5	.68	68	MSS SG	41	RM SG	75.0	102.3	45	DT GE	24	RM SG	57	QM SG	32
58	RP SG	160	179	162.5	.63	63	CE 5Y	41	RP SG	57.6	106.7	34	SS SF	21	RP SG	49	SS SF	31
59	SH SG	96	122	113.5	.36	36	STS GE	40	SH SG	56.6	91.1	3	EN SG	21	SH SG	20	MSS SG	31
....																		
61	SKS SG	41	59	43.5	.00	0	SH SG	36	SKS SG	63.4	87.8	0	YN SG	19	SKS SG	0	AS SG	28
62	SM SG	147	183	159.5	.41	41	EN AT	36	SM SG	49.8	84.6	0	AD SG	14	SM SG	21	AD SG	24

Distribution

AIR UNIVERSITY LIBRARY
ARMY MANAGEMENT STAFF COLLEGE LIBRARY
ARMY RESEARCH INSTITUTE LIBRARY
ARMY WAR COLLEGE LIBRARY
CENTER FOR NAVAL ANALYSES LIBRARY
DEFENSE TECHNICAL INFORMATION CENTER
HUMAN RESOURCES DIRECTORATE TECHNICAL LIBRARY
JOINT FORCES STAFF COLLEGE LIBRARY
MARINE CORPS UNIVERSITY LIBRARIES
NATIONAL DEFENSE UNIVERSITY LIBRARY
NAVAL HEALTH RESEARCH CENTER WILKINS BIOMEDICAL LIBRARY
NAVAL POSTGRADUATE SCHOOL DUDLEY KNOX LIBRARY
NAVAL RESEARCH LABORATORY RUTH HOOKER RESEARCH LIBRARY
NAVAL WAR COLLEGE LIBRARY
NAVY PERSONNEL RESEARCH, STUDIES, AND TECHNOLOGY SPISHOCK LIBRARY (3)
PENTAGON LIBRARY
USAF ACADEMY LIBRARY
US COAST GUARD ACADEMY LIBRARY
US MERCHANT MARINE ACADEMY BLAND LIBRARY
US MILITARY ACADEMY AT WEST POINT LIBRARY
US NAVAL ACADEMY NIMITZ LIBRARY